

AD _____

Award Number: W81XWH-07-1-0178

TITLE: Structural Basis for TSC-1 TSC-2 Complex Formation

PRINCIPAL INVESTIGATOR: John A.A. Ladas, M.D.

CONTRACTING ORGANIZATION: Beth Israel Deaconess Medical Center
Boston, MA 02215

REPORT DATE: Final

TYPE OF REPORT: March 2008

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE 01-03-2008		2. REPORT TYPE Final		3. DATES COVERED 15 Feb 2007– 14 Feb 2008	
4. TITLE AND SUBTITLE Structural Basis for TSC-1 TSC-2 Complex Formation				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER W81XWH-07-1-0178	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) John A.A. Ladas, M.D. Email: jladas@bidmc.harvard.edu				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Beth Israel Deaconess Medical Center Boston, MA 02215				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT Tuberous sclerosis complex (TSC) is a neurological disorder characterized by the formation of hamartomas in brain, skin, kidney, and other organs. The tumor suppressor genes TSC1 and TSC2 encode the proteins hamartin and tuberin, respectively. The tuberin(1-418) region interacts with hamartin(302-430), forming the TSC1–TSC2 complex that functions in cell growth regulation. Certain mutations in TSC patients disrupt the hamartin–tuberin interaction, indicating that association of these proteins is required for their growth suppressing function. However, the mechanisms underlying their interaction are poorly understood because their three-dimensional structures are not known. To obtain structural insights into the hamartin–tuberin interaction, the hamartin(302-430) and tuberin(11-418) domains were expressed in Escherichia coli cells and purified to homogeneity. Attempts to crystallize the isolated hamartin(302-430) and tuberin(11-418) domains yielded small crystals, whereas the hamartin–tuberin heterodimer did not produce crystals. Experiments are in progress to produce diffraction quality crystals of these proteins and determine their structures.					
15. SUBJECT TERMS Tuberous sclerosis complex, hamartin, tuberin, tumor suppressor gene, structural chemistry, X-ray crystallography					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	26	19b. TELEPHONE NUMBER (include area code)

Table of Contents

	<u>Page</u>
Introduction.....	4
Body.....	4
Key Research Accomplishments.....	7
Reportable Outcomes.....	7
Conclusion.....	7
References.....	8
Appendices.....	9

INTRODUCTION

Tuberous sclerosis complex (TSC) is a neurological disorder characterized by the formation of hamartomas in several organs, including brain, skin, kidney, heart, and liver (1-2). This syndrome often manifests in early age with infantile seizures and patients may have mental retardation and autism. Brain lesions include cerebral cortical tubers, subependymal nodules, and retinal hamartomas. Patients often develop skin lesions, including facial angiofibromas, hypomelanotic macules, and shagreen patches. In the kidney, benign angiomyolipomas occur frequently, whereas malignant angiomyolipomas and renal cell carcinomas are less common. Brain and renal lesions cause the highest morbidity and mortality among TSC patients (1-2).

The tumor suppressor genes *TSC1* and *TSC2* encode the proteins hamartin and tuberin, respectively (3,4), and their mutagenic inactivation occurs in a wide variety of hamartomas in TSC patients. Hamartin and tuberin are expressed in the same cell types within multiple organs including the kidney, brain, lung and pancreas, and interact directly forming a heterodimer (5). Hamartin is an 1164-amino acid hydrophilic protein expressed in most human tissues, and has a single transmembrane domain (amino acids 127-144) and a predicted coiled-coil region at residues 719-998. Tuberin is an 1807-amino acid protein and contains a GAP domain spanning residues 1517-1674 that inhibits the G protein Ras homolog enriched in brain (Rheb) (6,7). Rheb activates the mammalian target of Rapamycin (mTOR) and inactivates B-Raf (7). Activation of mTOR leads to increased protein synthesis. Upon insulin or growth factor stimulation, tuberin is phosphorylated by several kinases, including AKT/PKB, leading to inhibition of its GAP activity and activation of mTOR. Several unrelated patients with tuberous sclerosis have point mutations in the TSC2 GAP domain that influence the stability of the TSC1–TSC2 complex (8,9).

The tuberin region spanning residues 1-418 contains a hamartin-binding domain (Figure 1A), which is necessary to maintain hamartin in a soluble form in the cytosol (9). Within this region, amino acids 346–371 are predicted to form a putative coiled-coil, which is necessary but not sufficient to mediate the interaction with hamartin because additional N-terminal residues are also required for the efficient formation of the TSC1–TSC2 complex (9). The hamartin region spanning amino acids 302-430 mediates the interaction with tuberin (9). Importantly, certain non-truncating mutations in patients with TSC disrupt the hamartin–tuberin interaction (9). For example, non-truncating TSC-causing mutations located within the binding region of hamartin (N198_F199delinsI;593–595delACT) or tuberin (G294E and I365del), abolish or dramatically reduce the interaction of these proteins. In contrast, non-pathogenic missense polymorphisms of tuberin (R261W, M286V, R367Q) in the same region as the disease-causing TSC2 mutations do not affect this interaction, indicating that physical association of hamartin with tuberin is required for the growth suppressing functions of these proteins (9).

BODY

During the one-year award period we proceeded with the experiments proposed in the Tasks of the original application. A description of our progress in these studies follows below.

Task 1. To express, purify, and crystallize the hamartin and tuberin interaction domains. To crystallize the hamartin(302-430) and tuberin(1-418) domains it is necessary to produce large amounts (multi-milligram quantities) of these protein fragments in bacterial cells in soluble form and purified them to homogeneity. Toward this goal, the DNA fragments

encoding the human hamartin(302-430) and tuberin(1-418) regions were amplified by the polymerase chain reaction (PCR) using the human hamartin and tuberin cDNAs as templates, respectively, (kind gifts from Dr. Angela Hodges, MRC Centre for Neurodegeneration Research, King's College, London) and cloned them following standard protocols (10) into pET-6H2, a modified pET-16b prokaryotic vector (Novagen) that allows the expression of recombinant hexahistidine-tagged proteins in *Escherichia coli* BL21(DE3) cells, as we described previously (11). Numerous attempts to produce large amounts of soluble hamartin(302-430) and tuberin(1-418) proteins in several *E. coli* strains using this vector in a wide range of temperatures and in different culture media failed to produce large amounts of soluble proteins necessary for the crystallization experiments. The encountered difficulties to produce large amounts of soluble recombinant proteins delayed the successful completion of the proposed tasks within the one-year award period. To solve this problem, we used PCR-mediated site-directed mutagenesis to optimize many rare arginine codons in the open reading frames of hamartin(302-430) and tuberin(1-418) regions, and cloned the resulting DNA fragments into a series of prokaryotic expression vectors, including pGEX-2 (Pharmacia) to produce them as fusions with glutathione S-transferase, pET-43a (Novagen) to produce them as fusions with NusA, pMAL-c2 (New England BioLabs) to produce them as fusions with maltose-binding protein, and pTYB12 (NEB) to produce them as fusions with intein. The resulting constructs were verified by DNA sequencing and were used to transform *E. coli* BL21(DE3) cells. Extensive tests with these constructs resulted in the production of soluble hamartin(302-430) and tuberin(1-418) proteins. The hamartin(302-430) and tuberin(1-418) domains were purified using a combination of affinity, ion exchange, and size exclusion chromatography (Figure 1B,C). Crystallization experiments were performed with the pure proteins using sparse matrix screenings and vapor diffusion crystallization methods. Needle-like microcrystals of the hamartin(302-430) protein were obtained in several days at 20°C using the sitting drop vapor diffusion method (Figure 1D). These crystals are too small for analysis and experiments are in progress to optimize the crystallization conditions. Likewise, microcrystals of the tuberin(1-418) protein were obtained at 20°C using the sitting drop vapor diffusion method (Figure 1E). These crystals diffract to low resolution (Figure 1F) and we are in the process of optimizing the crystallization conditions in order to obtain larger crystals that diffract to higher resolution.

Task 2. To determine the crystal structures of the hamartin(302-430) and tuberin(1-418) domains. Because the hamartin(302-430) and tuberin(1-418) proteins interact to form a heterodimer, we assembled this complex by mixing equimolar amounts of these proteins and performed crystallization experiments. Despite numerous crystallization trials, no crystals of the hamartin(302-430)–tuberin(1-418) protein complex have been obtained to date. We will proceed with the crystallization of this complex by taking special care to purify monodisperse protein complexes for crystallization trials using size exclusion chromatography. For structure determination, crystals of selenomethionine-containing recombinant hamartin(302-430) and tuberin(1-418) proteins will be used to collect multiwavelength anomalous diffraction (MAD) diffraction data using synchrotron radiation. The crystal structures will be determined with MAD methods. **I would like to emphasize that although the award period is finished, we will continue to pursue very persistently the crystallographic analysis of the hamartin(302-430) and tuberin(1-418) proteins, as outlined in the original proposal.** The one-year Concept Award has provided critical support to initiate this research project and because of our progress in producing these proteins and initiating crystallization experiments, we plan to expand on these studies and apply this year to the Department of Defense for an Idea Development Award.

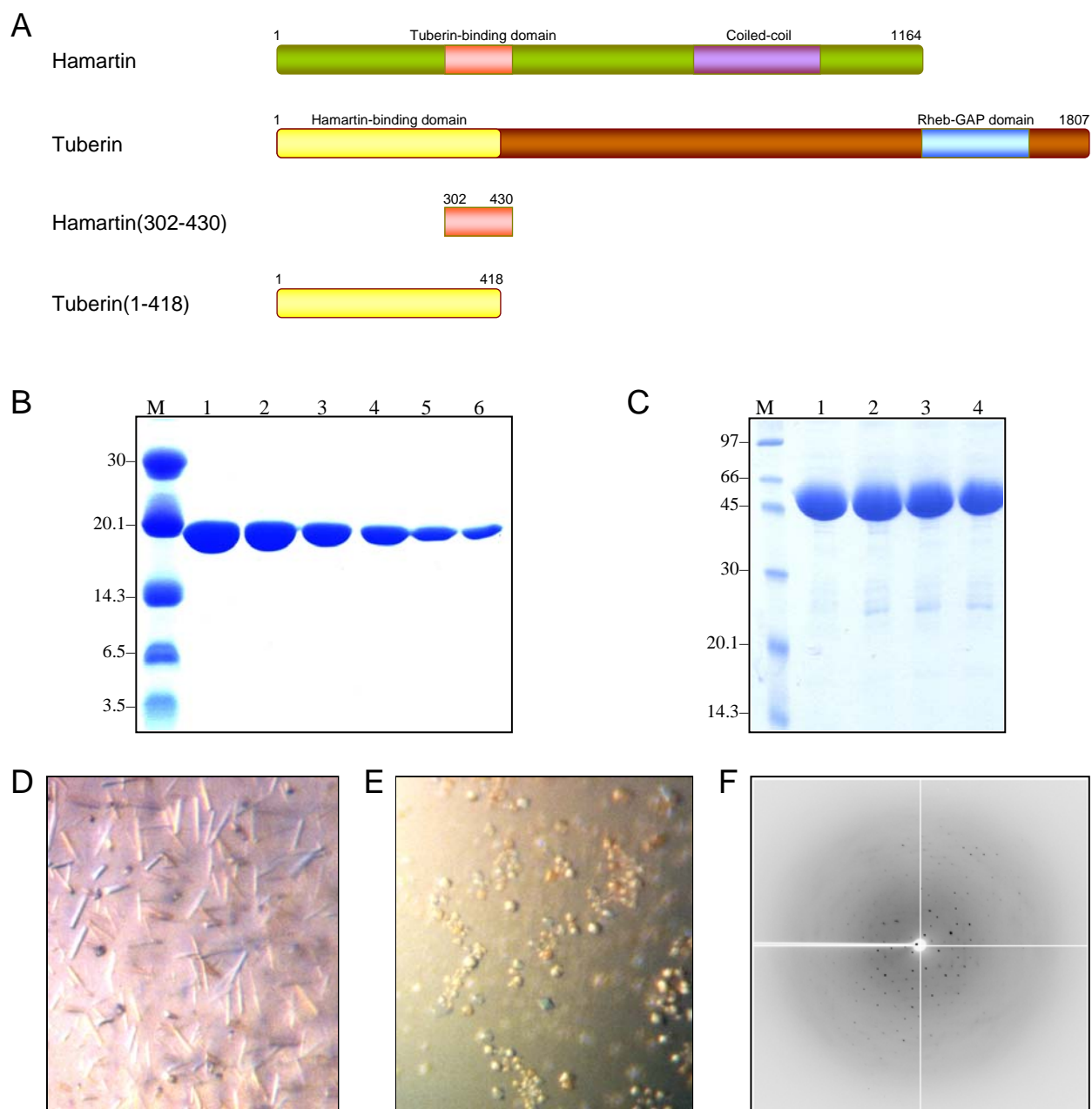


Figure 1. (A) Schematic representation of the human proteins hamartin and tuberin (not to scale) and their domains proposed to be analyzed crystallographically. Numbers indicate amino acid residues. (B) SDS-PAGE analysis of the hamartin(302-430) domain expressed in *E. coli* BL21(DE3) cells and purified using affinity, ion exchange, and size exclusion chromatography. Lane M: protein markers in kDa; lanes 1-6: elution fractions of purified hamartin(302-430) protein. (C) SDS-PAGE analysis of the tuberin(1-418) domain expressed in *E. coli* BL21(DE3) cells and purified using affinity chromatography. Lane M: protein markers in kDa; lanes 1-4: elution fractions of tuberin(1-418) protein. (D) Needle-like microcrystals of the hamartin(302-430) domain obtained by the sitting drop vapor diffusion method. (E) Microcrystals of the tuberin(1-418) domain obtained by the sitting drop vapor diffusion method. (F) Diffraction pattern of a tuberin(1-418) crystal collected on an R-Axis IV detector system with 3° oscillation frames exposed for 30 min at room temperature and crystal-to-detector distance of 150 mm.

KEY RESEARCH ACCOMPLISHMENTS

1. Cloning of the DNA fragment coding for the human hamartin(302-430) protein into prokaryotic expression vectors pET-6H2, pGEX-2, pET-43a, pMAL-c2, and pTYB12 for production in *E. coli* BL21(DE3) cells (*Task 1*).
2. Cloning of the DNA fragment coding for the human tuberin(1-418) protein into prokaryotic expression vectors pET-6H2, pGEX-2, pET-43a, pMAL-c2, and pTYB12 for production in *E. coli* BL21(DE3) cells (*Task 1*).
3. Induction of expression and performance of solubility tests of the bacterially produced human hamartin(302-430) protein (*Task 1*).
4. Induction of expression and performance of solubility tests of the bacterially produced human tuberin(1-418) protein (*Task 1*).
5. Large-scale production of the human hamartin(302-430) protein in *E. coli* cells and purification of this protein using a combination of affinity, ion exchange, and size exclusion chromatography (*Task 1*).
6. Large-scale production of the human tuberin(1-418) protein in *E. coli* cells and purification of this protein using affinity, ion exchange, and size exclusion chromatography (*Task 1*).
7. Sparse matrix crystallization experiments with the purified human hamartin(302-430) protein in the apo form, using the sitting drop vapor diffusion method (*Task 1*).
8. Sparse matrix crystallization experiments with the purified human tuberin(1-418) protein in the apo form, using the sitting drop vapor diffusion method (*Task 1*).
9. Sparse matrix crystallization experiments with the purified human hamartin(302-430) domain complexed with the tuberin(1-418) protein, using the sitting drop vapor diffusion method (*Task 1*).
10. Preliminary X-ray diffraction analysis of the tuberin(1-418) crystals with data collected on an R-Axis IV imaging plate detector system (*Task 2*).

REPORTABLE OUTCOMES

The experiments of the proposal are still in progress and no publications have resulted so far on the structural analysis of the hamartin and tuberin. Partial support provided by this award to our laboratory was acknowledged in the following two papers that we published during the award period:

1. Birrane G., Varma A.K., Soni A. and Ladas J.A.A. (2007). Crystal Structure of the BARD1 BRCT Domains. *Biochemistry* **46**: 7706-7712
2. Meiyappan M., Birrane G. and Ladas J.A.A. (2007). Structural Basis for Polyproline Recognition by the FE65 WW Domain. *Journal of Molecular Biology* **372**: 970-980

CONCLUSIONS

During the one-year award period we made numerous attempts to find conditions that would allow the production of large amounts of the human hamartin(302-430) and tuberin(1-418) protein fragments in *E. coli* cells, using the prokaryotic expression vectors pET-6H2, pGEX-2, pET-43a, pMAL-c2, and pTYB12. After solving several problems with the solubility of these

proteins, we were able to produce them in a soluble form appropriate for structural studies. We performed a large number of crystallization experiments with the hamartin(302-430) and tuberlin(1-418) proteins in the apo form and as a heterodimeric complex. Microcrystals of the unbound hamartin(302-430) and tuberlin(1-418) proteins were obtained using the sitting drop vapor diffusion method but the hamartin(302-430)–tuberlin(1-418) complex did not yield any crystals to date. We are currently in the process of optimizing the crystallization conditions of these proteins using fractional factorial experiments with the goal to obtain diffraction-quality crystals. We will continue to pursue these studies even after the expiration of the award period because the crystal structures of the hamartin(302-430) and tuberlin(1-418) proteins will elucidate the molecular mechanisms underlying the interaction and function of these proteins and will reveal the molecular changes induced by TSC-causing mutations in these proteins that abrogate their interaction and contribute to the pathogenesis of TSC.

REFERENCES

1. Jozwiak J. (2006). Hamartin and tuberlin: working together for tumour suppression. *Int. J. Cancer* **118**: 1-5
2. Astrinidis A, Henske EP. (2005). Tuberous sclerosis complex: linking growth and energy signaling pathways with human disease. *Oncogene* **24**: 7475-7481
3. European Tuberous Sclerosis Consortium. (1993). Identification and characterization of the tuberous sclerosis gene on chromosome 16. The European Chromosome 16 Tuberous Sclerosis Consortium. *Cell* **75**: 1305-1315
4. van Slegtenhorst M, de Hoogt R, Hermans C, Nellist M, Janssen B, Verhoef S, Lindhout D, van den Ouweland A, Halley D, Young J, Burley M, Jeremiah S, *et al.* (1997). Identification of the tuberous sclerosis gene TSC1 on chromosome 9q34. *Science* **277**: 805-808
5. van Slegtenhorst M, Nellist M, Nagelkerken B, Cheadle J, Snell R, van den Ouweland A, Reuser A, Sampson J, Halley D, van der Sluijs P. (1998). Interaction between hamartin and tuberlin, the TSC1 and TSC2 gene products. *Hum. Mol. Genet.* **7**: 1053-1057
6. Tee AR, Manning BD, Roux PP, Cantley LC, Blenis J. (2003). Tuberous sclerosis complex gene products, tuberlin and hamartin, control mTOR signaling by acting as a GTPase-activating protein complex toward Rheb. *Curr. Biol.* **13**: 1259-1268
7. Manning BD, Cantley LC. (2003). Rheb fills a GAP between TSC and TOR. *Trends Biochem. Sci.* **28**: 573-576
8. Nellist M, Verhaaf B, Goedbloed MA, Reuser AJ, van den Ouweland AM, Halley DJ. (2001). TSC2 missense mutations inhibit tuberlin phosphorylation and prevent formation of the tuberlin-hamartin complex. *Hum. Mol. Genet.* **10**: 2889-2898
9. Hodges AK, Li S, Maynard J, Parry L, Braverman R, Cheadle JP, DeClue JE, Sampson JR. (2001). Pathological mutations in TSC1 and TSC2 disrupt the interaction between hamartin and tuberlin. *Hum. Mol. Genet.* **10**: 2899-2905
10. Sambrook J, Fritsch EF, Maniatis T. (1989). *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press)
11. Hiremath CN, Ladias JAA. (1998). Expression and purification of recombinant hRPABC25, hRPABC17, and hRPABC14.4, three essential subunits of human RNA polymerases I, II, and III. *Prot. Expr. Purif.* **13**: 198-294

Crystal Structure of the BARD1 BRCT Domains^{†,‡}Gabriel Birrane, Ashok K. Varma,[§] Aditi Soni, and John A. A. Ladias**Molecular Medicine Laboratory and Macromolecular Crystallography Unit, Division of Experimental Medicine, Harvard Institutes of Medicine, Harvard Medical School, Boston, Massachusetts 02115**Received February 14, 2007; Revised Manuscript Received April 27, 2007*

ABSTRACT: The interaction of the breast tumor suppressor BRCA1 with the protein BARD1 results in the formation of a heterodimeric complex that has ubiquitin ligase activity and plays central roles in cell cycle checkpoint control and DNA repair. Both BRCA1 and BARD1 possess a pair of tandem BRCT domains that interact in a phosphorylation-dependent manner with target proteins. We determined the crystal structure of the human BARD1 BRCT repeats (residues 568–777) at 1.9 Å resolution. The composition and structure of the BARD1 phosphoserine-binding pocket P₁ are strikingly similar to those of the BRCA1 and MDC1 BRCT domains, suggesting a similar mode of interaction with the phosphate group of the ligand. By contrast, the BARD1 BRCT selectivity pocket P₂ exhibits distinct structural features, including two prominent histidine residues, His685 and His686, which may be important for ligand binding. The protonation state of these histidines has a marked effect on the calculated electrostatic potential in the vicinity of P₂, raising the possibility that ligand recognition may be regulated by changes in pH. Importantly, the BARD1 BRCT structure provides insights into the mechanisms by which the cancer-associated missense mutations C645R, V695L, and S761N may adversely affect the structure and function of BARD1.

The BRCA1-associated RING domain protein 1 (BARD1)¹ is the major binding partner of the breast and ovarian tumor suppressor BRCA1, and it plays a central role in the regulation of its stability, cellular localization, and function (1, 2). The BRCA1–BARD1 heterodimer has ubiquitin ligase activity and participates in the maintenance of genomic stability through its function in cell cycle checkpoint control and DNA repair (1, 2). In addition, BARD1 mediates apoptosis in a p53-dependent, BRCA1-independent manner, and BARD1-repressed cells are resistant to apoptosis (3, 4).

Human BARD1 is a 777 amino acid protein that contains a RING finger (residues 46–90), three ankyrin repeats (residues 420–525), and two tandem BRCT domains (residues 568–777) (5, 6). Germline and somatic mutations that affect the BARD1 structure occur in a subset of breast, ovarian, and uterine cancers, and some of these mutations have been identified in hereditary breast and ovarian cancers from patients negative for mutations in *BRCA1* (7–9). Two missense mutations (C557S and Q564H) occur in the region

between the ankyrin repeats and the BRCT domains, while others (N295S and K312N) occur outside the known domains. Three cancer-associated missense mutations, C645R, V695L, and S761N, occur within the BARD1 BRCT domains (Figure 1A). C645R has been associated with breast and ovarian cancer, V695L has been found in breast cancer, and S761N has been identified in uterine and breast cancer (7, 9). These mutations adversely affect the tumor suppressor function of BARD1 (9), but the underlying molecular mechanisms are not known.

BRCT repeats are phosphopeptide-binding modules found in a number of proteins that function in DNA-damage checkpoint control and DNA repair (10–12). Recent crystallographic studies of the BRCA1 and MDC1 BRCT domains bound to their target peptides offered mechanistic insights into ligand recognition by these modules (13–17). Specifically, a shallow pocket (P₁) in the N-terminal BRCT receives the phosphoserine (pSer) at position 0 of the ligand, whereas a deeper hydrophobic pocket (P₂) in the groove between the N- and C-terminal BRCT repeats selects residues at position +3. The amino acid composition and architecture of P₁ are similar among several BRCTs, consistent with its nondiscriminatory binding to pSer 0. In contrast, the P₂ pockets from various BRCT domains display significant differences that determine their selectivity properties. The molecular basis for the selection of Phe +3 by the BRCA1 BRCTs was revealed by the crystal structures of these domains bound to the BACH1 (13, 14) and CtIP (16) ligands, whereas the crystal structure of the MDC1 BRCTs bound to the C-terminal tail of γ -H2AX elucidated the mode of Tyr +3 recognition (17). However, oriented peptide library studies provided evidence that the BARD1 BRCTs select for ligands with the motif (pSer 0)-(Asp/Glu +1)-(Asp/Glu +2)-(Glu

[†] This work was supported by Grants GM065520, DK062162, and AG021964 from the National Institutes of Health, DAMD170210300, DAMD170310563, W81XWH0510622, and W81XWH0710178 from the U.S. Department of Defense, an Experienced Investigator Award from the Massachusetts Department of Public Health, and Temple Discovery Award TLL035927 from the Alzheimer's Association to J.A.A.L.

[‡] The atomic coordinates and structure factors have been deposited in the Protein Data Bank as entry 2NTE.

* Corresponding author. Tel: (617) 667-0064. Fax: (617) 975-5241. E-mail: jladias@bidmc.harvard.edu.

[§] Present address: Boston Biomedical Research Institute, Watertown, MA 02472.

¹ Abbreviations: BARD1, BRCA1-associated RING domain 1; BRCA1, breast cancer susceptibility gene 1; BRCT, BRCA1 C-terminal; MDC1, mediator of DNA damage checkpoint protein 1; SAD, single-wavelength anomalous dispersion; SeMet, selenomethionine.

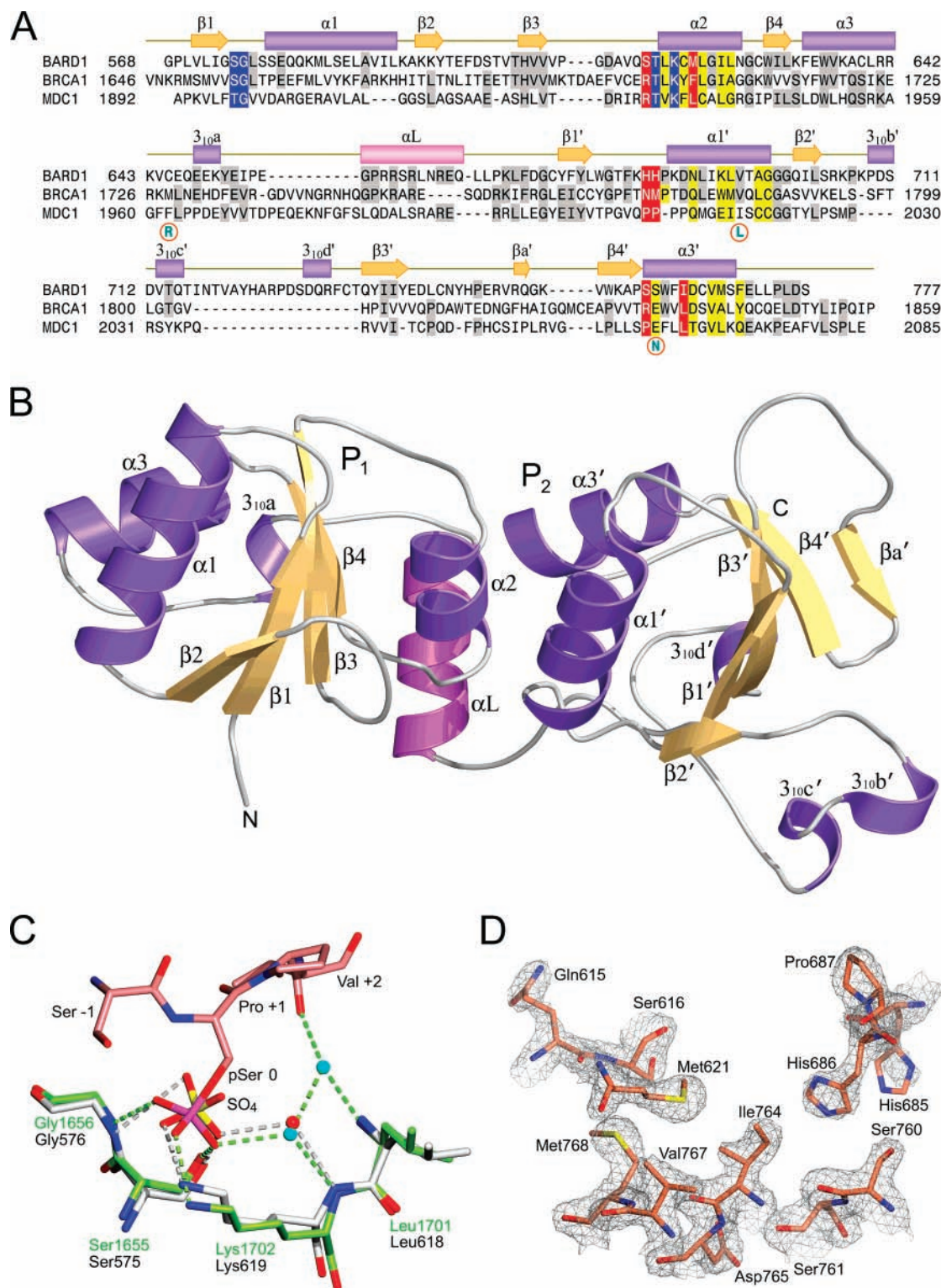


FIGURE 1: (A) Sequence alignment of the BARD1, BRCA1, and MDC1 BRCT domains. The protein sequences were aligned using the program CLUSTAL W (31) and manual intervention. Hyphens represent gaps inserted for optimum alignment. The secondary structure elements of BARD1, assigned by the program STRIDE (32), are depicted as orange arrows (β -strands) and purple cylinders (helices). The linker α -helix α L is colored pink. The BRCT2 secondary structure elements are labeled with primes. Residues lining the P_1 and P_2 pockets are shown as white letters on blue and red background, respectively. Identical amino acids in two or more sequences are shaded gray, while residues at the interface of the BRCT repeats are highlighted yellow. Cancer-associated missense mutations are encircled. (B) Ribbon representation of the BARD1 BRCTs. The location of the P_1 and P_2 pockets are denoted. Strands and helices are colored as in (A). The figure was made using BOBSCRIPT (33). (C) Stick model superposition of the P_1 pockets of the BRCA1–CtIP complex (PDB entry 1Y98) and the unbound BARD1. The CtIP peptide is colored pink, the interacting residues of BRCA1 are in green, and water molecules are cyan spheres. BARD1 residues are colored gray, and an associated water molecule is shown as a red sphere. Nitrogen, oxygen, phosphorus, and sulfur atoms are shown in blue, red, magenta, and yellow, respectively. Hydrogen bonds are dashed lines. The orientation of the P_1 pocket is identical to that of (B). (D) A weighted $2F_o - F_c$ electron density map of residues lining the P_2 pocket calculated at 1.9 Å resolution and contoured at 1.0σ . Atoms are colored as in (C). For clarity, the model has been rotated approximately 90° in relation to (B) such that the view is from underneath the P_2 pocket. The figure was made using POVSCRIPT (34) and POV-Ray (www.povray.org).

+3) (12), suggesting that the BARD1 selectivity determinants are different from those of other BRCT domains.

To elucidate the molecular mechanisms underlying the function of the BARD1 BRCT repeats, we have determined the crystal structure of these domains in the apo form. Comparison with the BRCA1 and MDC1 BRCTs reveals distinct structural differences in the C-terminal BRCT of BARD1, the most notable of which is the presence of His685 and His686 at the selectivity pocket P₂, suggesting their involvement in ligand recognition. In addition, the present work provides insights into the mechanisms by which cancer-associated missense mutations within the BRCT modules have detrimental effects on BARD1 function.

EXPERIMENTAL PROCEDURES

Protein Purification and Crystallization. A DNA fragment encoding the human BARD1 BRCT domains (residues 568–777) was amplified using the polymerase chain reaction and cloned into a modified pET-6H vector (18). The BARD1 BRCT protein carrying an N-terminal hexahistidine tag followed by a site for the human rhinovirus 3C protease (EVLFQ↓GP) was expressed in *Escherichia coli* BL21(DE3) cells grown at 37 °C until they reached an absorbance at 600 nm of 0.5, followed by induction with 0.5 mM IPTG at 20 °C for 20 h. The cells were harvested by centrifugation, resuspended in phosphate-buffered saline supplemented with protease inhibitor cocktail tablets (Roche Applied Science) on ice, and lysed on an EmulsiFlex-C3 homogenizer (Avestin). Although the majority of the recombinant BRCT protein was found in the insoluble fraction, an amount adequate for structural studies was recovered as soluble protein (~4 mg/L of culture). Soluble 6His-BRCT protein was purified on Ni-NTA resin (Qiagen), eluted with 200 mM imidazole, and dialyzed against a buffer containing 25 mM Tris-HCl (pH 7.0), 300 mM NaCl, and 2 mM DTT. The hexahistidine tag was removed by digestion with 3C protease, and the protein was further purified by size exclusion chromatography on Superdex 75 (GE Healthcare). The BARD1 BRCT protein was concentrated to 25 mg/mL by ultracentrifugation and crystallized in 20–25% polyethylene glycol 5000 mono-methyl ether, 150 mM ammonium sulfate, and 100 mM 2-morpholinoethanesulfonic acid (pH 6.5) at 20 °C by the sitting drop vapor diffusion method. Crystals were cryoprotected in mother liquor supplemented with 16% ethylene glycol and flash-frozen in a liquid nitrogen stream. BARD1 BRCT protein containing selenomethionine (SeMet) was expressed in B834(DE3) cells (Novagen) cultured in minimal media supplemented with SeMet (40 mg/L) and purified in a similar manner. Native and single-wavelength anomalous dispersion (SAD) data were collected on beamline X12B at the National Synchrotron Light Source (NSLS), Brookhaven National Laboratory, Upton, NY. The data were processed using HKL2000 (19). The crystals belong to space group $P2_12_12_1$ with unit cell dimensions $a = 57.4$ Å, $b = 75.8$ Å, $c = 116.6$ Å, and $\alpha = \beta = \gamma = 90^\circ$. There are two molecules in the asymmetric unit, referred to as A and B.

Structure Determination and Refinement. Data analysis, location of the selenium atoms, phase calculation, and density modification were carried out using SHELXC, SHELXD, and SHELXE (20), respectively. Phase extension and automated model building were performed with ARP/wARP

Table 1: Structure Determination and Refinement Statistics

data set	native	SAD
wavelength (Å)	0.9784	0.9796
resolution range (Å) ^a	50–1.90 (1.97–1.90)	50–2.04 (2.11–2.04)
no. of obsd reflections	309077	304610
no. of unique reflections	40185	32924
completeness (%) ^a	98.3 (91.7)	99.2 (98.9)
redundancy ^a	7.7 (6.1)	9.3 (8.9)
R_{sym} (%) ^b	6.4 (40.2)	8.7 (38.2)
overall $\langle I/\sigma(I) \rangle$	30.7 (4.3)	19.7 (5.5)
R_{cryst} (%) ^c	19.6	
R_{free} (%) ^d	24.7	
Ramachandran plot		
most favored (%)	91.0	
additionally allowed (%)	8.2	
generously allowed (%)	0.5	
disallowed (%)	0.3	
bond lengths ^e (Å)	0.022	
bond angles ^e (deg)	1.865	

^a Values in parentheses are for the highest resolution shell. ^b $R_{\text{sym}} = \sum |I - \langle I \rangle| / \sum I$, where I is the observed integrated intensity, $\langle I \rangle$ is the average integrated intensity obtained from multiple measurements, and the summation is over all observed reflections. ^c $R_{\text{cryst}} = \sum ||F_o| - k|F_c|| / \sum |F_o|$, where F_o and F_c are the observed and calculated structure factors, respectively. ^d R_{free} is calculated as R_{cryst} using 5% of the reflections chosen randomly and omitted from the refinement calculations. ^e Bond lengths and angles are root-mean-square deviations from ideal values.

(21), in combination with manual intervention using COOT (22). The final model, refined isotropically with eight TLS groups in REFMAC5 (23), contains 3382 protein atoms, 238 water molecules, 2 sulfate ions, 3 chloride ions, and 5 ethylene glycol molecules. Weak electron density in pocket P₂ was modeled as ethylene glycol in molecule A and as water in molecule B. The N-terminal residues Gly and Pro that remain after cleavage with 3C protease are native to BARD1 (Gly568 and Pro569). Gly568 of molecule B is disordered and was not modeled. PROCHECK (24) was used for analysis and validation of the refined structure. Glu655 of molecule A is at the edge of the disallowed region of the Ramachandran plot ($\phi = 75.7^\circ$, $\psi = -119.8^\circ$). The structure determination and refinement statistics are shown in Table 1.

Electrostatic Potential Calculation. The program GRASP (25) was used to calculate the electrostatic potentials. Charges were assigned as follows: Arg N^{η1} and Arg N^{η2}, +0.5 e; Glu O^{ε1}, Glu O^{ε2}, Asp O^{δ1}, and Asp O^{δ2}, −0.5 e; Lys N^ξ, 1.0 e. Dielectric constants of 2 and 80 were used for the interior protein and outer solvent regions, respectively. A charge of +0.5 e was used for N^{δ1} and N^{ε2} of the protonated His685 and His686, in addition to the above charges.

Isothermal Titration Calorimetry. The binding capability of BARD1 BRCT to various synthetic peptides was measured using a VP-ITC microcalorimeter (MicroCal). Briefly, 0.5 mM peptides SRSTpSDDENK and KEESpSDDDEDK were titrated against 0.05 mM BARD1 BRCT protein in phosphate-buffered saline (pH 7.4) supplemented with 150 mM NaCl at 23 °C. Similar experiments were also performed in 100 mM Tris-HCl (pH 7.4) and 150 mM NaCl and in 25 mM sodium phosphate (pH 6.0) and 400 mM NaCl at 23 °C. Titration curves were analyzed using the program ORIGIN 5.0 (OriginLab).

RESULTS AND DISCUSSION

Structure Determination. We expressed the human BARD1 BRCT domains (residues 568–777) carrying an N-terminal

hexahistidine tag followed by a 3C protease cleavage site in *E. coli* BL21(DE3) cells. We purified the soluble fraction of the recombinant protein using affinity chromatography, removed the tag by digestion with 3C protease, and crystallized the native protein using the sitting drop vapor diffusion method. The structure was determined by SAD, using synchrotron data from SeMet-substituted crystals. There are two molecules in the asymmetric unit, A and B, which display differences in the region spanning residues 708–718. An α -helix in this area in molecule B corresponds to two short 3_{10} -helices in molecule A. Since molecule A has superior electron density and its ligand-binding region is more solvent-exposed, thus resembling more closely the physiologic milieu, it will be used to describe the present structure. The final model was refined to a crystallographic factor R_{cryst} of 19.6% and an R_{free} of 24.7% with data extending to 1.9 Å resolution.

Structure of the BARD1 BRCT Repeats. The overall topology of the BARD1 BRCT domains is similar to previously determined BRCA1 and MDC1 BRCT structures. The region spanning residues 568–777 folds into two tandem domains, BRCT1 (residues 568–654) and BRCT2 (residues 669–777), linked by a central α -helix (α_L) (Figure 1B). BRCT1 comprises a central β -sheet formed by four parallel β -strands (β_1 – β_4) and flanked by the α -helices α_1 and α_3 on one side and α_2 on the other. BRCT2 also consists of a parallel β -sheet formed by the β -strands β_1' – β_4' , which are neighbored on one side by the α -helices α_1' and α_3' . However, the α_2' helix that exists between β_3' and β_4' in other BRCT2 structures is replaced by a short antiparallel β -strand (β_a') in BARD1 BRCT2 (Figure 1B). Another unique feature of the BARD1 BRCTs is the presence of three short 3_{10} -helices ($3_{10b'}$, $3_{10c'}$, and $3_{10d'}$) in the β_2' – β_3' loop.

The two BRCT modules pack closely against each other, burying a hydrophobic interface of ~ 1450 Å² and creating a surface groove with two pockets, P₁ and P₂. The hydrophilic P₁ is formed by residues Ser575, Gly576, Thr617, and Lys619 (Figure 1C), conforming to the consensus motif (Ser/Thr)-Gly...Thr-X-Lys that is characteristic of the pSer-interacting BRCA1 and MDC1 P₁ pockets (Figure 1A). A sulfate ion from the crystallization solution is present in the P₁ pocket, where it makes direct and solvent-mediated interactions with BRCT residues. Specifically, the sulfate ion hydrogen bonds with the main chain N of Gly576, the O' of Ser575, and the N ζ of Lys619 and via a water molecule to the main chain N of Lys619 (Figure 1C). These interactions are strikingly similar to those stabilizing the phosphate group of the ligand pSer 0 in the BRCA1 and MDC1 structures (13–17), suggesting that the BARD1 P₁ residues may also be involved in similar interactions with the phosphate group of the ligand.

The deeper and more hydrophobic pocket P₂ is lined by Ser616, Met621, His685, His686, and Ile764 (Figure 1D). Two prominent features of this pocket are His685 and His686, corresponding to Asn1774 and Met1775 of BRCA1, respectively, that mediate interactions with Phe +3 of the ligand (13–16), suggesting that the BARD1 histidines could play a key role in ligand selection. The electron density of the His685 imidazole is weak, indicating that this side chain is flexible (Figure 1D). By contrast, the imidazole group of His686 is well ordered, and its position and geometry at the rim of P₂ favor the formation of a hydrogen bond between

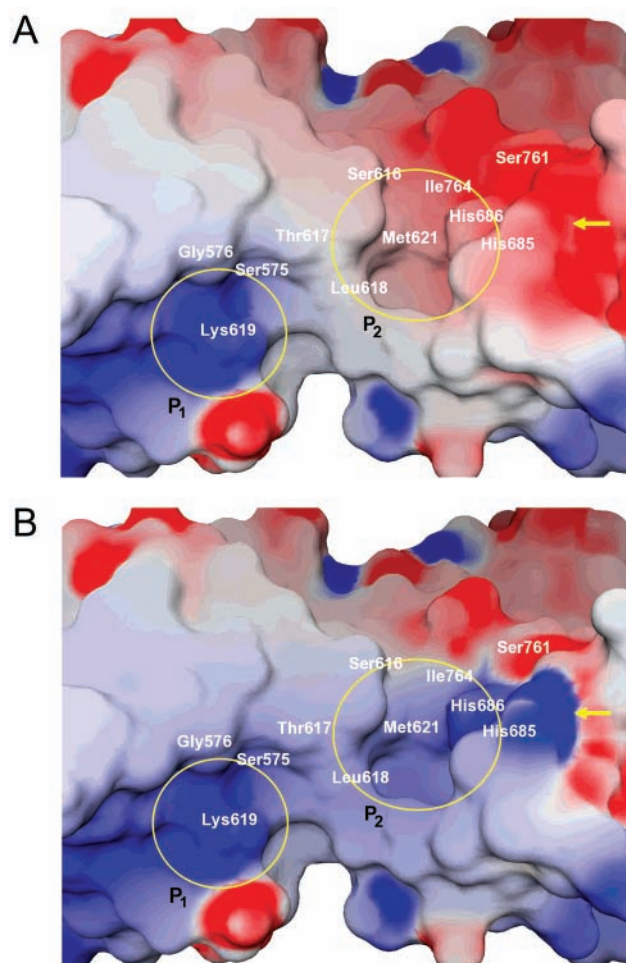


FIGURE 2: Electrostatic potential of the BARD1 BRCT surface (A) at near neutral pH and (B) at acidic pH. Pockets P₁ and P₂ are encircled, and the residues that form them are indicated. Arrows denote the change of the electrostatic potential due to the (de)protonation of His685 and His686. Electrostatic potentials were calculated with the program GRASP (25) and are colored as red (acidic, $-10k_B T$), white (neutral, $0k_B T$), and blue (basic, $10k_B T$).

its N^{ε2} atom and the side chain of a polar residue at position +3 of the ligand. Importantly, calculation of the electrostatic potential of the P₂ pocket shows that it is dramatically altered by changes in the protonation state of His685 and His686. At near neutral pH, the solvent-exposed N^{δ1} and N^{ε2} atoms of His685 and His686 are not protonated, and the P₂ pocket has a negative electrostatic potential (Figure 2A), whereas at more acidic pH, these atoms become protonated, switching the net charge to positive (Figure 2B). This raises the intriguing possibility that the BARD1 BRCT interaction with its ligand(s) may be dynamically regulated by the protonation of His685 and His686 in response to pH shifts in local cellular microenvironments during various physiological and pathological conditions (26–28). Notably, such a “histidine switch” mechanism regulates the interaction of the receptor-associated protein with the low-density lipoprotein receptor-related protein as a consequence of pH changes (29). Furthermore, it was recently reported that the protonation of a conserved histidine at the binding pocket of PSD-95 PDZ3 controls the interaction of this domain with its target peptide in a pH-dependent manner (30). It remains to be determined whether a similar mechanism is operational in ligand recognition by the BARD1 BRCT domains.

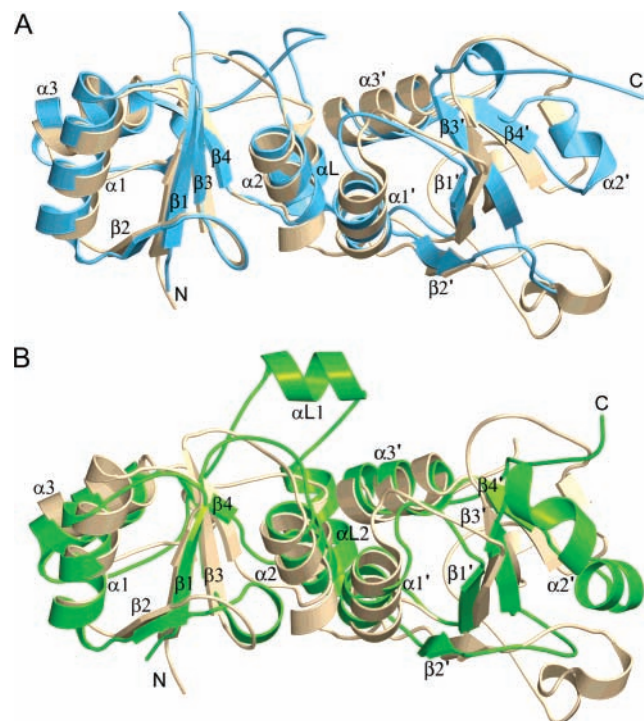


FIGURE 3: (A) Superposition of the crystal structures of the BARD1 BRCTs (beige) and the unbound BRCA1 BRCTs (light blue) (PDB entry 1JNX). (B) Superposition of the BARD1 BRCTs (beige) and the unbound MDC1 BRCTs (green) (PDB entry 2ADO). The models were aligned using secondary structure matching (35) in COOT.

Structural Comparison of the BARD1, BRCA1, and MDC1 BRCT Repeats. The BARD1 and BRCA1 BRCT structures are superimposed well, with a root-mean-square deviation of 1.7 Å for all C^α atoms (Figure 3A). However, there are distinct structural differences in these domains that likely reflect their functional differences. The β3–α2 loop of BARD1 is three residues shorter than its BRCA1 counterpart, and these regions are not superimposable. The BARD1 loop α3–αL is nine residues shorter and adopts an entirely different conformation from the corresponding loop of BRCA1. The β2'–β3' loop of BRCA1 is 19 residues shorter than that of BARD1 and is missing the three 3₁₀-helices of BARD1. Furthermore, the α2' helix in the β3'–β4' loop of BRCA1 is replaced by the βa' strand in BARD1.

The BARD1 and MDC1 BRCT structures are superimposed less well, with a root-mean-square deviation of 2.3 Å for all C^α atoms (Figure 3B). One notable difference is the presence of the α-helix αL1 in MDC1 that is absent in BARD1. Like BRCA1, the β2'–β3' loop of MDC1 is 21 residues shorter than that of BARD1 and is missing the three 3₁₀-helices of BARD1, whereas the BARD1 α3–αL loop is 10 residues shorter than that of MDC1. The αL–β1' loop of BARD1 is 2 residues longer than the corresponding loop of MDC1, resulting in poor superposition of these regions. Another difference is the presence of the elongated and bent α2' helix in the β3'–β4' loop of MDC1 that is replaced by the short βa' strand in BARD1.

Ligand Selection by the BARD1 BRCTs. To study the ligand-binding properties of the BARD1 BRCTs, we used isothermal titration calorimetry to measure the affinity of these domains for the synthetic peptides SRSTpSDDENK and KEESpSDDDEDK, which contain the consensus motif

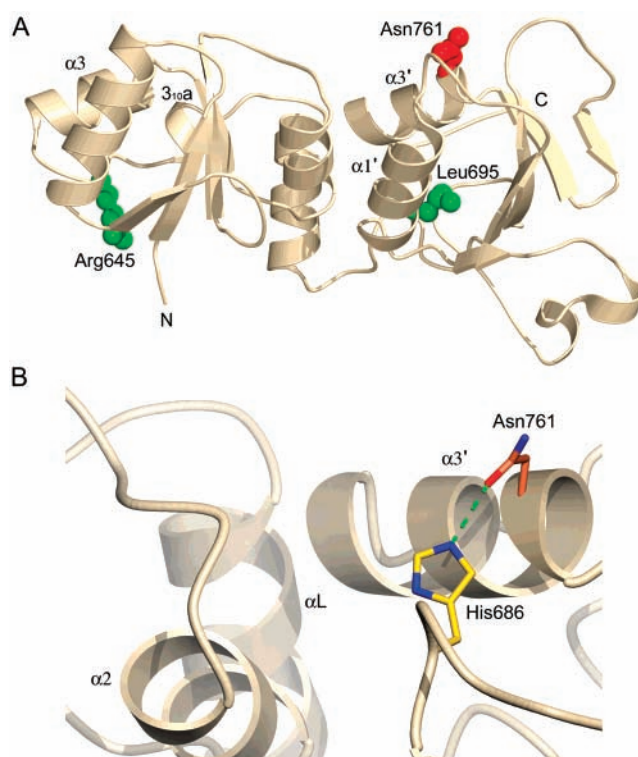


FIGURE 4: (A) Mapping of the cancer-associated missense mutations C645R, V695L, and S761N on the BARD1 BRCT structure. The side chains of Arg645 and Leu695 that are predicted to destabilize the BRCT fold are shown as green space-filled models, whereas the surface-exposed side chain of Asn761 is colored red. (B) Proposed mechanism for the abrogation of the BARD1 BRCT function by the S761N mutant. The Asn761 O^{δ1} is in close proximity to His686 N^{ε2} (2.7 Å) to form a hydrogen bond (green dotted line) that locks the imidazole ring of His686 in a conformation that cannot interact with the side chain of a residue at position +3 of the ligand. The figure was made using BOBSCRIPT and POV-Ray (www.povray.org).

(shown in bold) previously reported to be recognized by these domains in oriented library screens (12). Surprisingly, no measurable binding was detected with either of these peptides at pH 6.0 and 7.4 and under experimental conditions identical to those reported (12). Furthermore, cocrystallization of these peptides with the BARD1 BRCT domains yielded crystals of unbound BARD1 protein (data not shown). These results contradict the previously reported BARD1 BRCT selectivity (12) and could not be easily attributed to different experimental conditions. In this context, it is important to note that the observed BARD1 BRCT specificities for the ligand residues at selectivity positions 0 to +3 were relatively low (12), suggesting that the binding is very weak and perhaps transient. It is therefore possible that additional interactions between residues surrounding the ligand consensus motif and amino acids outside the P₁ and P₂ pockets are required for high-affinity binding. Alternatively, the BARD1 BRCT repeats may recognize macromolecular ligands with entirely different sequence motifs and structural characteristics. As part of our attempts to identify ligands for BARD1, we also tested the BACH1 peptide ISRSTpSPTFNKQ, which is an avid ligand for the BRCA1 BRCTs, and the doubly phosphorylated p53 peptide EPPLpSQEpTFS for binding to the BARD1 BRCT domains in similar isothermal titration calorimetry experiments. However, as expected from the nature of the BARD1 P₂ pocket, which is missing critical

hydrophobic residues that make contacts with the phenyl ring of Phe +3 in the BRCA1–ligand structures (13–16), no binding was observed with these peptides (data not shown). Therefore, the analysis of the structural determinants underlying ligand selection by the BARD1 BRCTs will await the identification of bona fide ligands for these domains.

Mapping of Cancer-Associated Missense Mutations. To obtain insights into the mechanisms underlying the disruption of BARD1 function by the C645R, V695L, and S761N mutations, we mapped the substituted residues on the BARD1 BRCT structure (Figure 4A). Cys645 is positioned in the loop between the $\alpha 3$ and 3_{10} helices. Substitution of this residue with a bulkier and charged arginine could disrupt the hydrophobic interface between $\beta 1$ and $\alpha 1$ formed by Leu570, Leu572, Leu592, and Ala594 and/or expose this interface by breaking hydrogen bonds between Glu648 O^{e2} and His606 N^{e2} on $\beta 3$ and between Trp635 N^{e1} and the carbonyl oxygen of Gln646, thereby destabilizing the BRCT1 fold. Val695 is part of the $\alpha 1'$ helix facing the β -sheet of BRCT2, and replacement of its isopropyl group by the bulkier isobutyl side chain of leucine could have a destabilizing effect on the BRCT2 fold. By contrast, Ser761 is surface-exposed and lies at the N-terminal end of $\alpha 3'$ in the vicinity of the BARD1 P₂ pocket. This residue corresponds to Glu1836 in the BRCA1–CtIP structure, which makes water-mediated hydrogen bonds with the carbonyl oxygen of Gly +4 of the peptide and the Arg1699 N⁷² of BRCA1 (16). Thus, it is possible that, in the S761N mutant, a hydrogen bond is formed between the Asn761 O^{δ1} and His686 N^{e2}, effectively locking the His686 imidazole ring in an unfavorable conformation to interact with the side chain of a residue at position +3 (Figure 4B), thereby reducing the BARD1 affinity for its binding partner.

Perspective. The crystal structure of the BARD1 BRCT domains reveals the atomic features of the pockets P₁ and P₂ and provides an essential basis for the elucidation of the mechanisms underlying ligand recognition by these modules during cell cycle control, DNA repair, and apoptosis. In light of our binding results, the previously reported selectivity of the BARD1 BRCTs for peptides having a glutamate at position +3 (12) needs to be independently verified and/or reevaluated by future studies aimed at dissecting the ligand specificity of these modules. When the physiologic binding partners of the BARD1 BRCTs are identified, it will be possible to study the role of His685 and His686 in ligand recognition and analyze the structural consequences of the cancer-associated missense mutations on the binding properties of these domains. Perhaps most importantly, the three-dimensional ligand-binding interface of the BARD1 BRCT repeats provides a structural framework for the design of small molecules that could specifically regulate the BARD1 function by modulating the BRCT affinities for their targets, with potential clinical applications.

ACKNOWLEDGMENT

We thank the staff at the NSLS for assistance during data collection and Dr. Donald Coen at Harvard Medical School for providing access to the microcalorimeter facility.

REFERENCES

- Irminger-Finger, I., and Jefford, C. E. (2006) Is there more to BARD1 than BRCA1? *Nat. Rev. Cancer* 6, 382–391.
- Narod, S. A., and Foulkes, W. D. (2004) BRCA1 and BRCA2: 1994 and beyond, *Nat. Rev. Cancer* 4, 665–676.
- Irminger-Finger, I., Leung, W. C., Li, J., Dubois-Dauphin, M., Harb, J., Feki, A., Jefford, C. E., Soriano, J. V., Jaconi, M., Montesano, R., and Krause, K. H. (2001) Identification of BARD1 as mediator between proapoptotic stress and p53-dependent apoptosis, *Mol. Cell* 8, 1255–1266.
- Feki, A., Jefford, C. E., Berardi, P., Wu, J. Y., Cartier, L., Krause, K. H., and Irminger-Finger, I. (2005) BARD1 induces apoptosis by catalysing phosphorylation of p53 by DNA-damage response kinase, *Oncogene* 24, 3726–3736.
- Wu, L. C., Wang, Z. W., Tsan, J. T., Spillman, M. A., Phung, A., Xu, X. L., Yang, M. C., Hwang, L. Y., Bowcock, A. M., and Baer, R. (1996) Identification of a RING protein that can interact in vivo with the BRCA1 gene product, *Nat. Genet.* 14, 430–440.
- Ayi, T. C., Tsan, J. T., Hwang, L. Y., Bowcock, A. M., and Baer, R. (1998) Conservation of function and primary structure in the BRCA1-associated RING domain (BARD1) protein, *Oncogene* 17, 2143–2148.
- Thai, T. H., Du, F., Tsan, J. T., Jin, Y., Phung, A., Spillman, M. A., Massa, H. F., Muller, C. Y., Ashfaq, R., Mathis, J. M., Miller, D. S., Trask, B. J., Baer, R., and Bowcock, A. M. (1998) Mutations in the BRCA1-associated RING domain (BARD1) gene in primary breast, ovarian and uterine cancers, *Hum. Mol. Genet.* 7, 195–202.
- Ghimenti, C., Sensi, E., Prescittini, S., Brunetti, I. M., Conte, P., Bevilacqua, G., and Caligo, M. A. (2002) Germline mutations of the BRCA1-associated ring domain (BARD1) gene in breast and breast/ovarian families negative for BRCA1 and BRCA2 alterations, *Genes Chromosomes Cancer* 33, 235–242.
- Sauer, M. K., and Andrulis, I. L. (2005) Identification and characterization of missense alterations in the BRCA1 associated RING domain (BARD1) gene in breast and ovarian cancer, *J. Med. Genet.* 42, 633–638.
- Manke, I. A., Lowery, D. M., Nguyen, A., and Yaffe, M. B. (2003) BRCT repeats as phosphopeptide-binding modules involved in protein targeting, *Science* 302, 636–639.
- Yu, X., Chini, C. C., He, M., Mer, G., and Chen, J. (2003) The BRCT domain is a phospho-protein binding domain, *Science* 302, 639–642.
- Rodriguez, M., Yu, X., Chen, J., and Songyang, Z. (2003) Phosphopeptide binding specificities of BRCA1 COOH-terminal (BRCT) domains, *J. Biol. Chem.* 278, 52914–52918.
- Shiozaki, E. N., Gu, L., Yan, N., and Shi, Y. (2004) Structure of the BRCT repeats of BRCA1 bound to a BACH1 phosphopeptide: implications for signaling, *Mol. Cell* 14, 405–412.
- Clapperton, J. A., Manke, I. A., Lowery, D. M., Ho, T., Haire, L. F., Yaffe, M. B., and Smerdon, S. J. (2004) Structure and mechanism of BRCA1 BRCT domain recognition of phosphorylated BACH1 with implications for cancer, *Nat. Struct. Mol. Biol.* 11, 512–518.
- Williams, R. S., Lee, M. S., Hau, D. D., and Glover, J. N. (2004) Structural basis of phosphopeptide recognition by the BRCT domain of BRCA1, *Nat. Struct. Mol. Biol.* 11, 519–525.
- Varma, A. K., Brown, R. S., Birrane, G., and Ladas, J. A. A. (2005) Structural basis for cell cycle checkpoint control by the BRCA1–CtIP complex, *Biochemistry* 44, 10941–10946.
- Stucki, M., Clapperton, J. A., Mohammad, D., Yaffe, M. B., Smerdon, S. J., and Jackson, S. P. (2005) MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks, *Cell* 123, 1213–1226.
- Hiremath, C. M., and Ladas, J. A. A. (1998) Expression and purification of recombinant hRPABC25, hRPABC17, and hRPABC14.4, three essential subunits of human RNA polymerase I, II and III, *Protein Expression Purif.* 13, 198–204.
- Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode, *Methods Enzymol.* 276, 307–326.
- Sheldrick, G., and Schneider, T. (1997) SHELXL: High-resolution refinement, *Methods Enzymol.* 277, 319–343.
- Perrakis, A., Morris, R., and Lamzin, V. S. (1999) Automated protein model building combined with iterative structure refinement, *Nat. Struct. Biol.* 6, 458–463.
- Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Crystallogr. D* 60, 2126–2132.
- Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method, *Acta Crystallogr. D* 53, 240–255.

24. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr.* **26**, 283–291.
25. Nicholls, A., Sharp, K. A., and Honig, B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons, *Proteins* **11**, 281–296.
26. Boron, W. F. (2004) Regulation of intracellular pH, *Adv. Physiol. Educ.* **28**, 160–179.
27. Boyer, M. J., and Tannock, I. F. (1992) Regulation of intracellular pH in tumor cell lines: influence of microenvironmental conditions, *Cancer Res.* **52**, 4441–4447.
28. Chesler, M. (2005) Failure and function of intracellular pH regulation in acute hypoxic-ischemic injury of astrocytes, *Glia* **50**, 398–406.
29. Lee, D., Walsh, J. D., Mikhailenko, I., Yu, P., Migliorini, M., Wu, Y., Krueger, S., Curtis, J. E., Harris, B., Lockett, S., Blacklow, S. C., Strickland, D. K., and Wang, Y. X. (2006) RAP uses a histidine switch to regulate its interaction with LRP in the ER and Golgi, *Mol. Cell* **22**, 423–430.
30. Chi, C. N., Engstrom, A., Gianni, S., Larsson, M., and Jemth, P. (2006) Two conserved residues govern the salt and pH dependencies of the binding reaction of a PDZ domain, *J. Biol. Chem.* **281**, 36811–36818.
31. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Res.* **22**, 4673–4680.
32. Frishman, D., and Argos, P. (1995) Knowledge-based protein secondary structure assignment, *Proteins* **23**, 566–579.
33. Esnouf, R. M. (1997) An extensively modified version of MolScript that includes greatly enhanced coloring capabilities, *J. Mol. Graphics Modell.* **15**, 132–134.
34. Fenn, T. D., Ringe, D., and Petsko, G. A. (2003) *Povscript+*: a program for model and data visualization using persistence of vision ray-tracing, *J. Appl. Crystallogr.* **36**, 944–947.
35. Krissinel, E., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions, *Acta Crystallogr. D60*, 2256–2268.

BI700323T



Structural Basis for Polyproline Recognition by the FE65 WW Domain

Muthuraman Meiyappan, Gabriel Birrane and John A. A. Ladas*

Molecular Medicine Laboratory
and Macromolecular
Crystallography Unit, Division
of Experimental Medicine
Harvard Institutes of Medicine
Harvard Medical School
Boston, MA 02115, USA

The neuronal protein FE65 functions in brain development and amyloid precursor protein (APP) signaling through its interaction with the mammalian enabled (Mena) protein and APP, respectively. The recognition of short polyproline sequences in Mena by the FE65 WW domain has a central role in axon guidance and neuronal positioning in the developing brain. We have determined the crystal structures of the human FE65 WW domain (residues 253–289) in the apo form and bound to the peptides PPPPPPLPP and PPPPPPPPL, which correspond to human Mena residues 313–321 and 347–356, respectively. The FE65 WW domain contains two parallel ligand-binding grooves, XP (formed by residues Y269 and W280) and XP2 (formed by Y269 and W271). Both Mena peptides adopt a polyproline helical II conformation and bind to the WW domain in a forward (N–C) orientation through selection of the PPPPP motif by the XP and XP2 grooves. This mode of ligand recognition is strikingly similar to polyproline interaction with SH3 domains. Importantly, comparison of the FE65 WW structures in the apo and liganded forms shows that the XP2 groove is formed by an induced-fit mechanism that involves movements of the W271 and Y269 side-chains upon ligand binding. These structures elucidate the molecular determinants underlying polyproline ligand selection by the FE65 WW domain and provide a framework for the design of small molecules that would interfere with FE65 WW–ligand interaction and modulate neuronal development and APP signaling.

© 2007 Elsevier Ltd. All rights reserved.

*Corresponding author

Keywords: FE65; WW domain; Mena; polyproline helix II; crystal structure

Introduction

FE65 is expressed predominantly in the brain, where it plays critical roles in cortical development and APP signaling.¹ FE65 possesses a WW domain (FE65 WW) and two phosphotyrosine-binding (PTB) domains with distinct binding specificities. The FE65 WW recognizes polyproline sequences present in several proteins, including the c-Abl tyrosine kinase,² the mammalian homolog of the

Drosophila protein enabled (Mena),³ and the neuronal P2X₂ receptor.⁴ The FE65 PTB1 binds to the histone acetyltransferase TIP60,⁵ the transcription factor CP2,⁶ and the cytoplasmic domains of the low-density lipoprotein receptor-related protein⁷ and ApoER2 receptor,⁸ whereas PTB2 interacts with the APP intracellular domain.⁹ The multi-protein complexes resulting from FE65-mediated interactions underlie the function of this protein in regulation of neuronal growth cone motility,^{10,11} ATP-mediated synaptic transmission,⁴ learning and memory,¹² APP processing,^{13–15} and transcriptional regulation of gene expression.^{5,16–22} FE65 WW is essential for many of these functions. In particular, the FE65 WW-mediated interaction with Mena plays a central role in axonal growth cone dynamics, axon guidance, and neuronal positioning in the developing brain.^{23–27}

WW domains, named for two conserved tryptophan residues (Figure 1(a)), are highly compact protein–protein interaction modules that fold into

Present address: M. Meiyappan, Xtal BioStructures, 313 Pleasant Street, Watertown, MA 02472, USA.

Abbreviations used: APP, amyloid precursor protein; GST, glutathione-S-transferase; MAD, multiwavelength anomalous dispersion; Mena, mammalian enabled; PPII, polyproline helix II; SAD, single-wavelength anomalous dispersion; SH3, Src homology 3.

E-mail address of the corresponding author:
johnladas@gmail.com

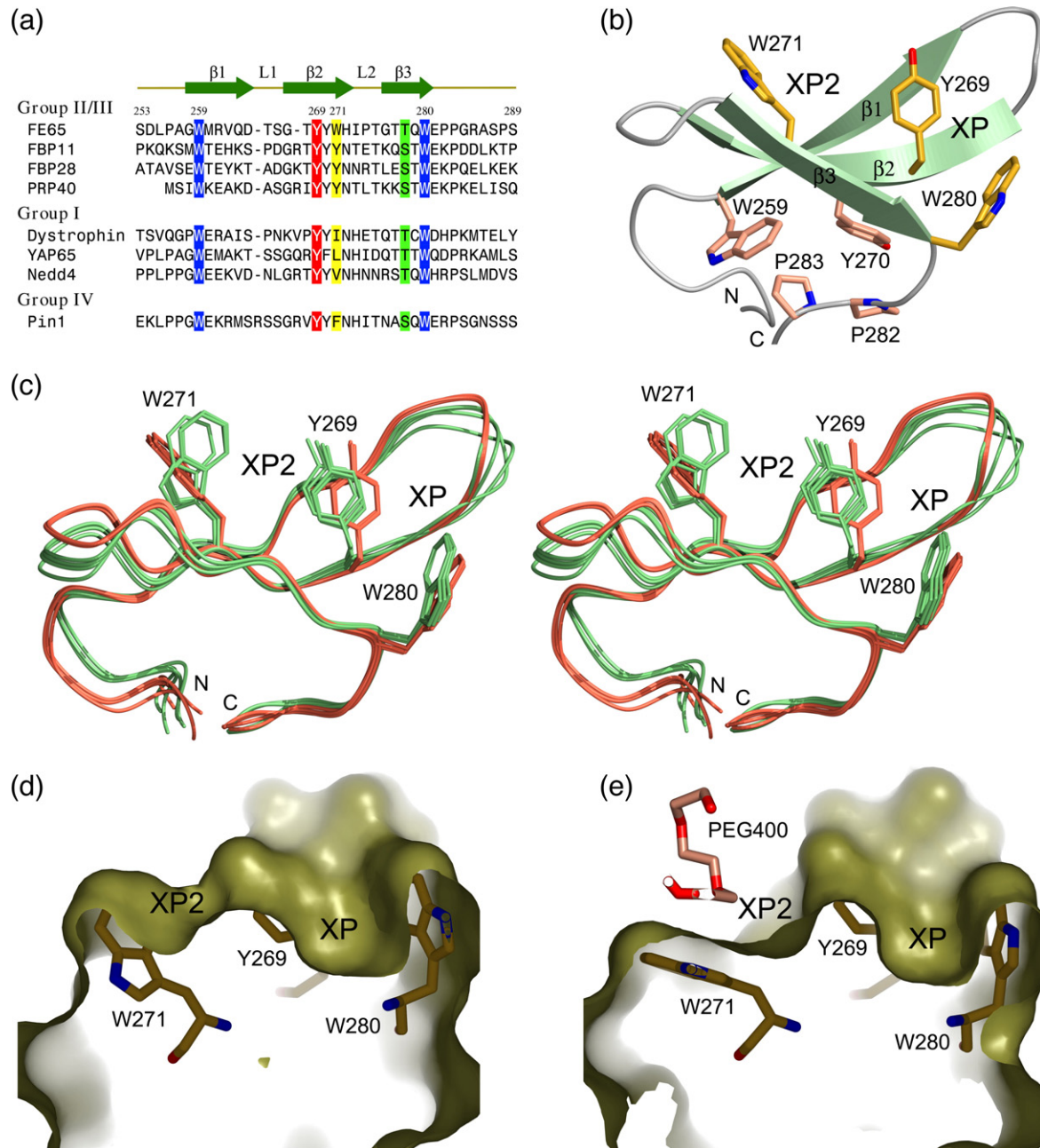


Figure 1. (a) Sequence comparison of representative WW domains. The sequences of the single WW of FE65, dystrophin, and Pin1, the first WW of FBP11, PRP40, YAP65, and Nedd4, and the second WW of FBP28, were aligned using the program CLUSTAL W.⁵⁶ Hyphens represent gaps inserted for optimum alignment. The secondary structure elements of FE65 WW are indicated at the top. Residues of human FE65 are numbered. Two conserved tryptophan residues (after which the domain is named) are shown in white on a blue background. Invariant tyrosine residues (shown in white on a red background) and bulky hydrophobic residues (highlighted in yellow) form the XP2 groove in group II/III domains. Conserved threonine/serine residues that hydrogen bond with the ligand are highlighted in green. (b) Ribbon diagram of the FE65 WW domain. Residues that form the XP and XP2 grooves are shown as yellow stick models. The side-chains of residues forming a hydrophobic core that stabilizes the fold are shown in pink. (c) A stereo view of eight FE65 WW molecules present in the asymmetric unit. Superposition of these independent structures shows that they fall into two distinct conformations highlighted in shades of green (apo form) and pink (bound to a PEG400 molecule). Note the conformational changes of the W271 and Y269 side-chains in the bound form. (d) In the apo form, the XP2 groove is shallow. (e) Binding of a PEG400 molecule induces the formation of a deep XP2 groove, primarily through conformational changes of the indole side-chain of W271 and, to a lesser degree, of the phenyl ring of Y269. The Figure was made using PyMol [www.pymol.org].

stable, slightly curved, three-stranded antiparallel β -sheet structures.^{28–31} These domains interact with short proline-rich sequences and are found in pro-

teins of diverse functions, including regulation of transcription, RNA processing, ubiquitin ligation, protein trafficking, receptor signaling, and control of

the cytoskeleton.^{28,32} WW domains were initially categorized into four groups on the basis of their ligand-binding specificity. Group I binds polypeptides with the core consensus PPXY (PY motif), where X can be any residue; group II recognizes the sequence PPLP (PL motif); group III selects for proline-rich sequences with arginine residues (PR motif); and group IV interacts with phospho(serine/threonine)-proline containing peptides (pS/pTP motif).²⁸ The concave surface of the WW domain contains a series of exposed, nearly parallel aromatic residues, forming a hydrophobic cavity that receives the polyproline peptide. Group I and IV domains have a single groove formed by two highly conserved aromatic residues (typically tyrosine in the β 2 strand and tryptophan in the β 3 strand), which receives the XP residues of the peptide (XP groove).^{28,29} By contrast, group II and III WW domains have an additional groove, XP2, which is formed by the conserved tyrosine and an aromatic residue in the β 2 strand.^{33,34} In this respect, groups II and III are more similar to SH3 domains, which have two successive XP grooves, suggesting that these structurally and evolutionarily unrelated protein modules have converged on a similar mechanism for polyproline recognition. On the basis of the distinct structural features and similar binding specificities of group II and III WW domains, including recognition of uninterrupted polyproline sequences (PPPPP) in addition to PL and PR motifs, it was proposed recently that these groups be merged into a larger class II/III.³⁵

Structural studies of WW domains bound to their cognate ligands, including the crystal structures of dystrophin³⁶ (group I) and Pin1³⁷ (group IV), and the solution structures of YAP65,³⁸ Nedd4,^{39,40} and Smurf2⁴¹ (group I), and FBP11^{33,34} (group II/III) WW–peptide complexes, provided insights into the molecular mechanisms underlying ligand selectivity by these modules. The peptides adopt a polyproline II (PPII) helical conformation with core proline residues packing against the XP and XP2 grooves, and carbonyl oxygen atoms hydrogen bonding with WW residues. Interactions between residues outside the ligand core motif and variable loops of the WW domain, as well as adjacent domains, contribute to the specificity and affinity.^{29,30,41} Importantly, WW domains can bind to ligands in opposite orientations, as was first observed with SH3 domains.^{42,43} For example, the Pin1 WW binds to the ligand in an N–C or + orientation,³⁷ whereas the dystrophin WW interacts with the peptide in a C–N or – orientation,³⁶ because of a twofold rotational pseudosymmetry in the shapes and hydrogen bonding networks of the WW grooves and PPII helices.²⁹

Insights into the architecture of the XP and XP2 grooves of group II/III domains were obtained from the solution structure of FBP11 WW1 bound to a peptide having the core sequence PPLP.^{33,34} A prominent feature of this structure is the non-parallel arrangement of the XP2 and XP grooves, with the former being occupied by the pair PP and the latter receiving the residues LP. However, because this is

currently the only structure of a group II/III WW domain, it is not clear whether the observed orientation of the XP and XP2 grooves is a common feature of all members of this group. The FE65 WW in particular, is distinct from other group II/III domains because it has a tryptophan (W271) instead of tyrosine in the β 2 strand that forms the XP2 groove (Figure 1(a)), suggesting that the size and orientation of this groove might be different. Therefore, structures of the FE65 WW and other group II/III WW domains bound to their ligands will advance our understanding of the molecular basis underlying ligand recognition by these protein modules.

Here, we report the atomic structures of FE65 WW in the apo form and bound to two polyproline peptides from human Mena, which represent the first high-resolution crystallographic analysis of a group II/III WW domain. The FE65 WW contains two parallel ligand-binding grooves, XP and XP2, the latter being formed by an induced-fit mechanism. These grooves select for the PPPPP motif of the Mena peptides using a mechanism strikingly similar to that of polyproline recognition by SH3 domains. These structures will facilitate the design of small molecules that could modulate the function of FE65, with possible clinical applications.

Results and Discussion

Structure determination

We expressed a protein fragment of human FE65 (residues 253–289) that encompasses the WW domain, as a glutathione-S-transferase (GST) fusion in *Escherichia coli* cells, purified it by affinity chromatography, and released the FE65 WW from the GST moiety by digestion with the tobacco etch virus (TEV) protease. The FE65 WW protein was further purified by size-exclusion chromatography. The purified protein formed multimeric complexes in solution, as judged by size-exclusion chromatography and dynamic light-scattering experiments (data not shown), and it was crystallized by the sitting-drop, vapor-diffusion method. Initial attempts to determine the crystal structure of the apo FE65 WW by molecular replacement using other WW structures as search models failed (data not shown). The structure was determined by single-wavelength anomalous dispersion (SAD) methods of selenomethionine (SeMet)-substituted crystals. The structure was refined to 2.28 Å resolution with a crystallographic factor R_{cryst} of 21.6% and an R_{free} of 28.2% (Table 1). Interestingly, there are eight molecules (A–H) in the asymmetric unit that adopt two distinct conformations (see below). We also crystallized FE65 WW bound to the peptide Mn10 (PPPPPPPPPL), corresponding to residues 347–356 of human Mena, and determined the structure by multiwavelength anomalous dispersion (MAD) methods, using crystals of Mn10 bound to SeMet-substituted FE65 WW. The crystal structure was refined to 1.33 Å resolution with an R_{cryst} of

Table 1. Statistics of structure determination and refinement

	WW Apo ^a		WW–Mn10 ^b				WW–Mn9 ^b
<i>A. Data collection and phasing</i>							
Data set	Native	SAD λ 1	Native	MAD λ 1	MAD λ 2	MAD λ 3	Native
Wavelength (Å)	0.975	0.9789	0.985	0.9792	0.9793	0.960	1.000
Resolution (Å)	50–2.19	50–2.53	50–1.33	50–1.58	50–1.58	50–1.50	50–1.27
Unique reflections	20584	13311	9129	5591	5595	6494	14379
Completeness (%) ^c	99.4(98.6)	98.9(89.2)	98.6(90.3)	99.9(100)	99.9(100)	99.3(93.2)	97.6(93.5)
Redundancy (%)	9.2(8.2)	36.0(27.1)	6.5(5.4)	20.5(19.6)	19.9(19.1)	12.8(9.6)	7.5(5.4)
$R_{\text{sym}}^{\text{d}}$ (%)	4.5(32.3)	8.5(42.7)	5.7(44.0)	7.1(31.3)	6.3(32.8)	6.5(43.8)	4.4(38.6)
$\langle I \rangle / \langle \sigma(I) \rangle$	43.3(5.3)	60.7(7.8)	30.1(3.8)	53.3(12.9)	52.6(12.1)	41.3(5.3)	27.5(3.7)
<i>B. Refinement</i>							
Resolution range (Å)	37.3–2.28		17.03–1.33				25.8–1.35
Reflections in working set	17,415		8654				11,423
Reflections in test set	924		431				586
$R_{\text{cryst}}^{\text{e}}$ (%)	21.6		17.2				19.2
$R_{\text{free}}^{\text{e}}$ (%)	28.2		21.4				22.4
Average isotropic B factors							
Protein (Å ²)	47.2		18.9				23.6
Peptide (Å ²)			20.3				29.6
Ramachandran plot							
Allowed regions (%)	87.8		91.3				92.9
Additionally allowed regions (%)	12.2		8.7				3.6
Generously allowed regions (%)	0.0		0.0				3.6
Disallowed regions (%)	0.0		0.0				0.0
Number of water molecules	119		34				36

^a Crystals of FE65 WW in the apo form were analyzed at NSLS beamline X12C.

^b Crystals of the FE65 WW–Mn10 and FE65 WW–Mn9 complexes were analyzed at NSLS beamline X12B.

^c The numbers in parentheses correspond to the highest resolution shell (2.27–2.19 Å for WW Apo Native, 2.62–2.53 Å for SAD, 1.38–1.33 Å for WW–Mn10 Native, 1.64–1.58 Å for MAD λ 1 and λ 2, 1.55–1.50 Å for MAD λ 3, and 1.32–1.27 Å for WW–Mn9 Native).

^d $R_{\text{sym}} = \sum |I - \langle I \rangle| / \sum I$, where I is the observed integrated intensity, $\langle I \rangle$ is the average integrated intensity obtained from multiple measurements, and the summation is over all observed reflections.

^e $R_{\text{cryst}} = \sum ||F_{\text{obs}}| - k|F_{\text{calc}}| / \sum |F_{\text{obs}}|$, where F_{obs} and F_{calc} are the observed and calculated structure factors, respectively.

^f R_{free} is calculated as R_{cryst} using 5% of the reflection data chosen randomly and omitted from the refinement calculations.

17.2% and an R_{free} of 21.4%. Using this structure as a search model, we also solved by molecular replacement the crystal structure of the FE65 WW bound to the peptide Mn9 (PPPPPLPP) spanning Mena residues 313–321, and refined it to 1.35 Å resolution with an R_{cryst} of 19.2% and an R_{free} of 22.4%. Both FE65 WW–Mn9 and FE65 WW–Mn10 crystals have one molecule in the asymmetric unit.

Crystal structure of FE65 WW in the apo form

FE65 WW adopts the typical WW fold, comprising an antiparallel β -sheet formed by three β -strands (β 1– β 3) and two short loops (L1 and L2) (Figure 1(a) and (b)). At the convex surface of the β -sheet the conserved aromatic residues Y270 and W259 stack against P282 and the invariant P283, respectively, forming a hydrophobic cluster that stabilizes the fold (Figure 1(b)). This hydrophobic core is shielded from the solvent by L255 and P256, and it brings together the N and C termini, generating a domain that can be transferred as a functional unit between proteins. The concave ligand-binding surface contains two distinct grooves, XP and XP2. The canonical XP groove is formed by Y269 and W280, and its size is determined by the L1 loop, which is essentially a short β -turn, and T278 in the β 3 strand. The XP2 groove is formed primarily by Y269, W271, and V262. In loop L2, H272 stacks against P274 exposing the hydrophobic side-chain of I273, which

in turn restricts the movement of the W271 indole ring and indirectly influences the XP2 groove. Because FE65 WW has only 20 residues between the conserved W259 and W280 (Figure 1(a)), its H272 has shifted towards the N terminus by one residue compared to other WW domains and, as a result, its imidazole ring lies below the β -sheet. By contrast, the corresponding histidine in group I WW domains lies above the β -sheet and participates in ligand recognition by hydrogen bonding with the tyrosine of the PY motif.^{36,38–41}

Formation of the FE65 WW XP2 groove by an induced-fit mechanism

Superimposition of the eight molecules present in the asymmetric unit shows that they are grouped into two distinct conformations (Figure 1(c)). Molecules B, C, E, and H adopt a conformation in which the XP2 groove is unoccupied and the side-chains of W271 and Y269 are approaching each other, effectively closing this groove (Figure 1(c) and (d)). By contrast, molecules A, D, F and G have a PEG400 molecule (from the crystallization solution) bound to the XP2 groove, resulting in a rotation of the W271 indole ring plane by an average of 60° relative to the unbound one (Figure 1(c) and (e)). The phenyl ring of Y269 is also displaced by ~1.5 Å, effectively increasing the width of the XP2 groove to accommodate the ligand. Sequence alignment shows that

W271 corresponds to smaller aromatic residues (Tyr or Phe) in other group II/III and IV domains and hydrophobic residues in group I domains (Figure 1(a)). Notably, mutation of the corresponding leucine (L190) to tryptophan in the YAP65 WW1 domain shifts the specificity of this domain from group I to that of group II.⁴⁴ The present crystal structure provides an explanation for this switch and elucidates the molecular basis for the important role of the W271 indole ring in the formation of the XP2 groove by an induced-fit mechanism.

Structural determinants of polyproline recognition by FE65 WW

It was shown that the human Mena polyproline sequences spanning residues 313–321 and 347–356 are ligands for the FE65 WW.³ To evaluate the contribution of these motifs to the Mena interaction with FE65 WW, we synthesized the peptides Mn9 and Mn10 that correspond to these regions and determined their binding affinities for FE65 WW using isothermal titration calorimetry. FE65 WW binds to Mn9 and Mn10 peptides with dissociation constants $124(\pm 9.5)$ μ M and $116(\pm 6.5)$ μ M, respectively (Figure 2). The micromolar affinities are in agreement with previously measured affinities for other WW–ligand complexes,⁴⁵ and are thought to underlie the formation of transient complexes by these modules during dynamic cellular processes.

The structural basis for the specificity and affinity of the FE65 WW–Mn10 interaction was elucidated

by the crystal structure of this complex (Figure 3(a) and (b)). The peptide adopts a PPII helix, and each of the XP and XP2 grooves is occupied by two proline residues: P5'/P6' insert into XP2 and P8'/P9' occupy XP (peptide residues denoted with primes). The residues P1'–P4' extend beyond the WW molecule (Figure 3(a)), with P1'/P2' contacting loop L2 of an adjacent crystallographic molecule. Notably, the side-chain of L10' faces toward the solvent and does not make any contact with the protein (Figure 3(a)). The XP2 groove is in the open conformation, as observed in the PEG400-bound FE65 WW. Indeed, the W271 and Y269 side-chains of FE65 WW bound to Mn10 superimpose almost exactly with the corresponding side-chains of the PEG400-bound FE65 WW, indicating that an induced-fit mechanism underlies polyproline ligand recognition by FE65 WW. Importantly, the separation of the two grooves by Y269 and T278 necessitates the presence of a spacer residue between the two pairs of proline residues that occupy the grooves. Although in the FE65 WW–Mn10 complex the spacer is proline (P7'), this position could be occupied by any amino acid (X) without breaking the PPII conformation, indicating that the sequence PPXPP is the general recognition motif of FE65 WW.

Because the Mn10 peptide does not contain an intact PPLP motif, reported to be recognized by FE65 WW,³ we also determined the crystal structure of FE65 WW bound to the peptide Mn9 that contains this sequence. Surprisingly, the XP and XP2 grooves of FE65 WW also bind to the PPPPP and not to the

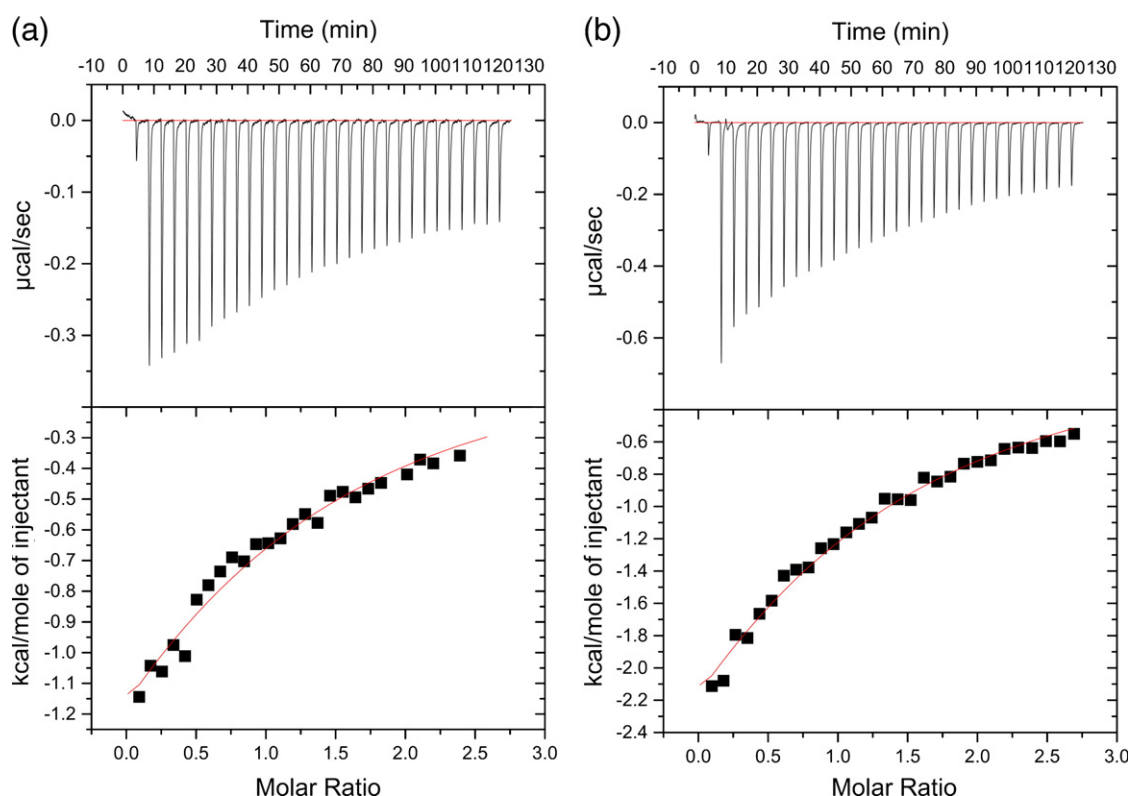


Figure 2. Isothermal titration calorimetry results obtained for the FE65 WW interaction with (a) peptide Mn10 and (b) peptide Mn9, as described in Experimental Procedures.

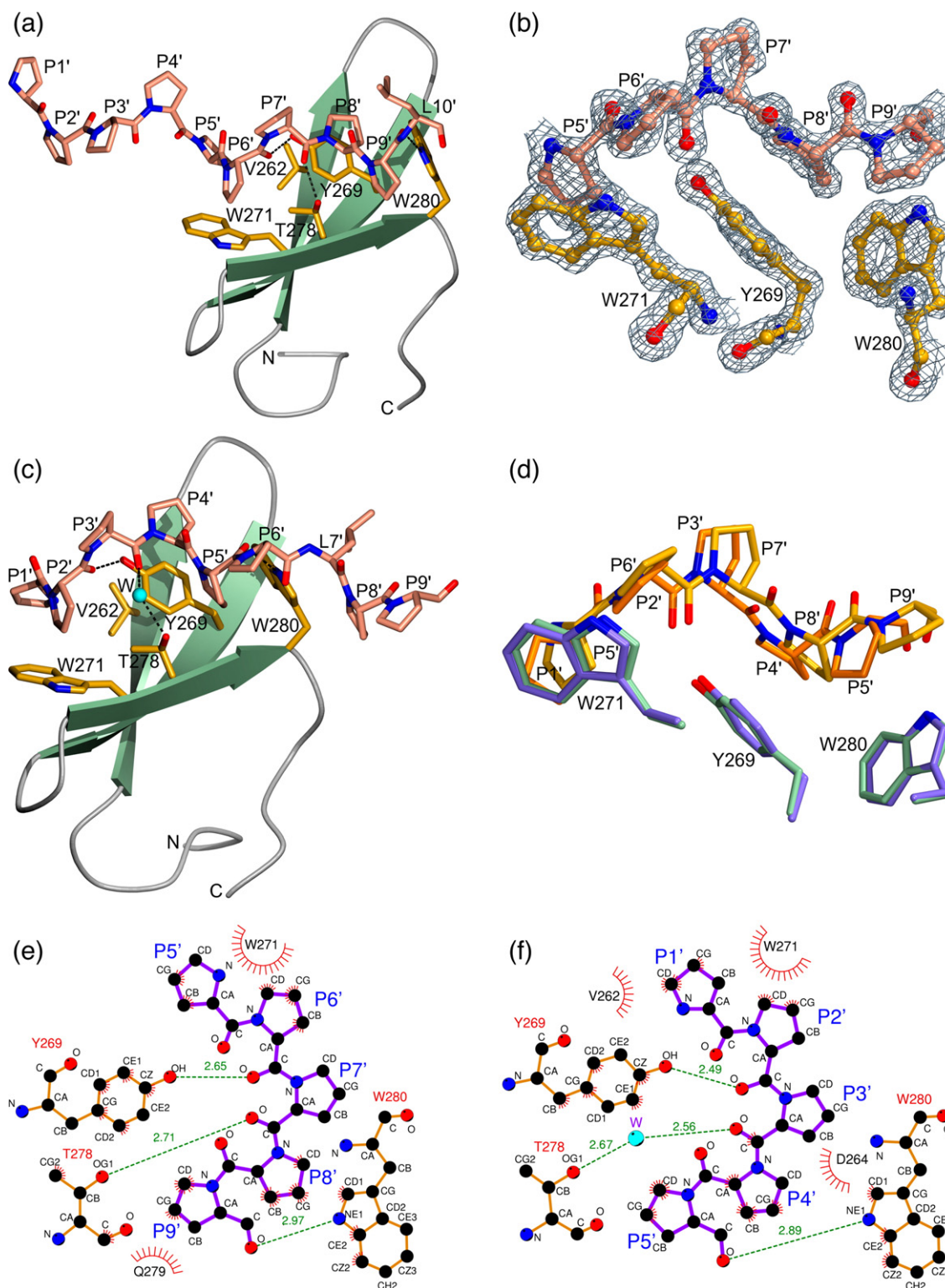


Figure 3. (a) A ribbon diagram of the FE65 WW bound to peptide Mn10, shown as a stick model. WW residues critical for interaction with the peptide are shown as yellow stick models. N and O atoms are colored blue and red, respectively. Hydrogen bonds are denoted by broken lines. (b) A weighted $2F_{\text{obs}} - F_{\text{calc}}$ electron density map of the FE65 WW-Mn10 structure calculated at 1.33 Å and contoured at 2σ . (c) A ribbon diagram of the FE65 WW bound to peptide Mn9. Atoms are colored as in (a). A water molecule is indicated by a cyan sphere. (d) Superposition of the FE65 WW (green) bound to Mn10 (yellow) with the FE65 WW (cyan) bound to Mn9 (orange) structures. For clarity, only the parts of the ligands that interact with the XP and XP2-forming aromatic residues of FE65 WW are shown. The Figure was made using BOBSRIPT⁵⁷ and POV-Ray (www.povray.org). (e) and (f) Two-dimensional representation of the FE65 WW interactions with the Mn10 and Mn9 peptides, respectively (made using LIGPLOT⁵⁸). FE65 and Mena residues are colored orange and purple, respectively. Hydrogen bonds are shown as broken lines, hydrophobic interactions as arcs with radial spokes, and a water molecule as a cyan sphere.

PPLP motif (Figure 3(c)). Like Mn10, Mn9 forms a PPII helix, with P1'P2' residues inserting into XP2 and P4'P5' occupying the XP groove, further corroborating the notion that the FE65 grooves have a stronger preference for proline rather than leucine. It appears that the small size of the FE65 XP groove, which is determined by the short loop L1, plays an important role in the selection of the pyrrolidine ring of proline over the bulkier isobutyl group of a leucine residue. By contrast, the longer L1 loop of FBP11 WW1 generates a larger XP groove that can accommodate a leucine residue. Nevertheless, in the FE65 WW–Mn9 structure the side-chain of L7' clamps the indole group of W280 outside the XP groove, providing additional stabilization to the complex (Figure 3(c)).

Superposition of the FE65 WW–Mn10 and FE65 WW–Mn9 complexes shows that the overall structures are very similar (root-mean-square deviation 0.61 Å for all C $^{\alpha}$ atoms of the WW region spanning L255–P283 and the atoms of the five proline residues that occupy the two grooves). However, it is noted that the Mn10 and Mn9 proline residues occupying the XP groove are not superimposed as well as those inserting into the XP2 pocket (Figure 3(d)), likely reflecting the different crystal packing arrangements of the two complexes due to differences in the crystallization conditions and/or peptide composition. Notably, the C-terminal region of the Mn9 peptide extends beyond the WW domain and the side-chain of L7' makes hydrophobic contacts with adjacent crystallographic molecules, resulting in a more tightly coiled PPII helix. Notwithstanding the possible crystal lattice effects that may influence the position of the peptide inside the XP and XP2 grooves, these structures demonstrate the flexibility of the FE65 WW grooves to accommodate slightly different positions of the ligand (Figure 3(d)). Both the FE65 WW complexes with the Mn10 and Mn9 peptides are stabilized by similar interactions, although subtle differences are observed in the hydrophobic contacts and hydrogen bonding networks of the two structures (Figure 3(e) and (f)). Most notably, the carbonyl oxygen atoms of P6' and P9' in Mn10 (P2' and P5' in Mn9) form direct hydrogen bonds with the hydroxyl group of Y269 and the N $^{\epsilon 1}$ of W280, respectively. By contrast, the O $^{\gamma 1}$ atom of T278 hydrogen bonds directly with the carbonyl oxygen atom of the spacer residue P7' in the FE65 WW–Mn10 complex but indirectly through a water molecule with the carbonyl oxygen atom of P3' in the FE65 WW–Mn9 structure.

Similarities between the FE65 WW and SH3 ligand-binding surfaces

A prominent feature of the FE65 WW–Mn10 and FE65 WW–Mn9 structures is the parallel orientation of the XP and XP2 grooves, which is due in large part to the W271 indole ring that forms the external wall of XP2 (Figure 4(a) and (b)). By comparison, two solution structures of the FBP11 WW1 bound to the peptide PPLP revealed that the XP and XP2

grooves are not parallel (Figure 4(c)).^{33,34} The FBP11 WW1 XP2 groove is formed by two tyrosine residues (Y23 and Y25) instead of a tyrosine and a tryptophan in FE65 (Y269 and W271), and the FBP11 grooves are nearly continuous because of the smaller side-chain of S32 compared to the T278 in FE65 WW. As a result, the FBP11 WW1 grooves are occupied by successive pairs of residues, without a spacer amino acid. In contrast, the parallel and non-continuous nature of the FE65 WW grooves underlies the recognition of a ligand having two pairs of proline residues separated by one spacer amino acid, the side-chain of which makes no contributions to the interaction.

The parallel orientation of the FE65 WW grooves is a novel structural feature of group II/III domains and is remarkably similar to the arrangement of the SH3 proline-binding grooves (Figure 4(a) and (d)), despite the different folds of these protein modules. Indeed, superposition of the FE65 WW–Mn10 and the Abl SH3–3BP1 crystal structures shows strikingly similar positions of the three aromatic residues, which emanate from different secondary structure elements in the WW and SH3 domains to form the two polyproline-binding grooves (Figure 4(e)). The almost identical spatial position of these critical aromatic triads provides strong evidence that these evolutionarily unrelated domains have converged upon a similar structural mechanism for PPII recognition.^{29,46}

Perspective

This work presents the first crystal structure of a group II/III WW domain and provides mechanistic insights into uninterrupted polyproline ligand recognition by FE65 WW at the atomic level. A prominent feature of this domain is the presence of a second groove, XP2, which is parallel with the conserved groove XP and is formed by an induced-fit mechanism. The parallel and non-continuous arrangement of the XP and XP2 grooves necessitates a spacer residue between the two pairs of proline residues, leading to the recognition of the PPXPP sequence, where X is any residue. Remarkably, the architecture of these grooves is more similar to that of SH3 domains than FBP11 WW1, the only other group II/III WW domain with known structure, providing evidence for evolutionary convergence. Thus, the high-resolution crystal structures of FE65 WW bound to polyproline ligands advances our understanding of the principles underlying ligand selection by this versatile domain. This information will facilitate the design of novel WW domains with desired ligand specificities and affinities, as described.^{47,48} Importantly, the specific and low-affinity interactions of WW domains with their polyproline ligands make them excellent targets for drug development.⁴⁵ Given the significance of the FE65 WW-mediated protein interactions in brain development and pathogenesis of human diseases, including Alzheimer's disease, the present work will provide a framework for the development of small

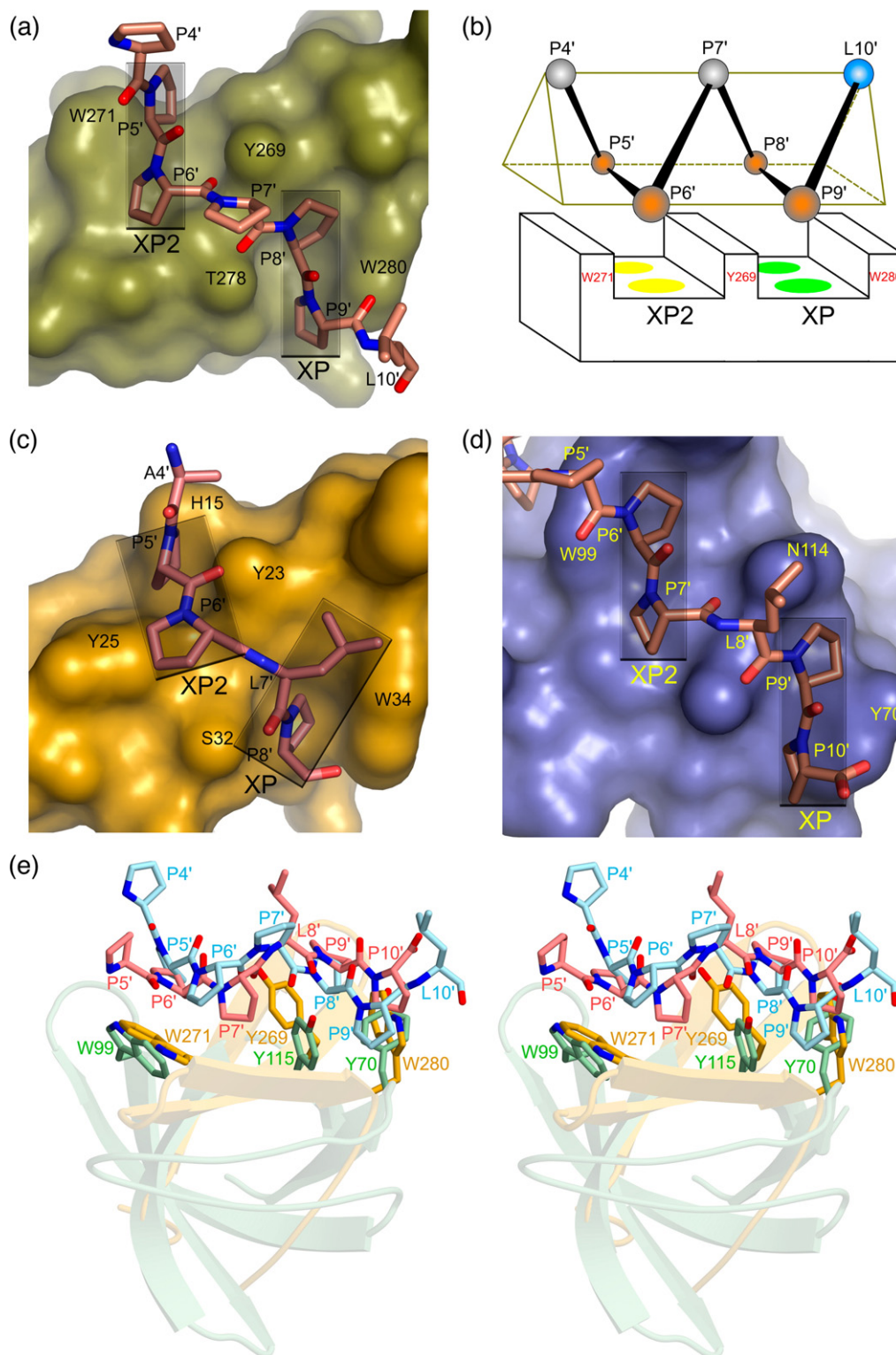


Figure 4. (a) Surface representation of the FE65 WW bound to the Mn10 peptide (shown as a stick model). Semitransparent rectangles denote the parallel orientation of the XP and XP2 grooves. (b) A diagram of the FE65 WW bound to Mn10 PPII helix (shown as a triangular prism), illustrating the parallel arrangement of the XP and XP2 grooves, the three aromatic residues that form them, and the proline residues of the ligand that occupy them (orange spheres). Unbound proline and leucine residues are shown as gray and cyan spheres, respectively. (c) Surface representation of the FBP11 WW1 bound to PPLP sequence (PDB code 2DYF), showing the non-parallel arrangement of the XP and XP2 grooves (rectangles). (d) Surface representation of the Abl tyrosine kinase SH3 domain bound to the peptide 3BP1 containing the sequence PPLPP (PDB code 1ABO), showing the parallel XP and XP2 grooves (rectangles). (e) A stereo view of the FE65 WW–Mn10 structure superimposed on the Abl SH3–3BP1 complex, showing the similar mode of PPII recognition by the aromatic triad of WW (yellow) and SH3 (light green) domains. The Mn10 and 3BP1 peptides are shown in cyan and pink, respectively.

molecules that could modulate these interactions and regulate the function of FE65, with possible clinical applications.

Experimental Procedures

Protein expression and crystallization

A DNA fragment encoding the human FE65 WW (residues 253–289) with a recognition site for the TEV protease (ENLYFQ↓G) at its N terminus was amplified by the polymerase chain reaction method and was cloned into a modified pGEX-2T vector (GE Healthcare). The FE65 WW protein was expressed in *E. coli* BL21(DE3) cells as a GST fusion, purified on glutathione Sepharose, released from GST by digestion with TEV, and was further purified by size-exclusion chromatography on a Superdex 30 column (GE Healthcare). SeMet-substituted FE65 WW protein was produced in *E. coli* B834 cells (Novagen) grown in M9 medium supplemented with 40 mg/l of SeMet (Sigma), and the protein was purified in the same manner as the unlabelled protein in the presence of 5 mM DTT throughout the entire purification process.

The FE65 WW protein was concentrated by ultracentrifugation to 40 mg/ml in 20 mM Tris–HCl (pH 7.6), 150 mM NaCl, and was used for crystallization experiments by the sitting-drop, vapor-diffusion method. Crystals of FE65 WW in the apo form were obtained in 0.1 M Hepes (pH 7.5), 2.2 M ammonium sulfate, 2% (w/v) PEG400 at 20 °C. Crystals were cryoprotected by stepwise transfer in 0.1 M Hepes (pH 7.5), 2.3 M ammonium sulfate, 10% (v/v) glycerol, 10% (w/v) xylitol. The crystals belong to space group $P6_322$ with unit cell dimensions $a=b=75.61$ Å, $c=226.49$ Å, $\alpha=\beta=90^\circ$, $\gamma=120^\circ$. There are eight molecules in the asymmetric unit. The crystallized FE65 WW protein includes the vector-derived residue G at its N terminus.

For crystallization of FE65 WW–Mena complexes, the protein concentration was adjusted to 18 mg/ml in 20 mM Tris–HCl (pH 7.6), 150 mM NaCl, mixed with the peptides Mn10 or Mn9 at ~1:3 (protein/peptide) molar ratios, incubated at 4 °C for 1–2 h, and used for crystallization by the sitting-drop, vapor-diffusion method. Crystals of the FE65 WW–Mn10 complex grew in 0.1 M Hepes (pH 7.4), 1.7–1.9 M ammonium sulfate, at 20 °C, and they were cryoprotected in 0.1 M Hepes (pH 7.4), 1.8 M ammonium sulfate, 20% glycerol. They belong to space group $P3_121$ with unit cell dimensions $a=b=41.66$ Å, $c=38.63$ Å, $\alpha=\beta=90^\circ$, $\gamma=120^\circ$, and have one molecule in the asymmetric unit. Crystals of the FE65 WW–Mn9 complex grew in 0.1 M Hepes (pH 7.6), 1.3 M sodium citrate, and were cryoprotected in 0.05 M Hepes (pH 7.6), 1.52 M sodium citrate. They belong to space group $P4_12_12$ with unit cell dimensions $a=b=37.50$ Å, $c=78.46$ Å, $\alpha=\beta=\gamma=90^\circ$, and there is one molecule in the asymmetric unit. The data were processed using the HKL2000 package.⁴⁹ Data collection and processing statistics are given in Table 1.

Structure determination and refinement

The crystal structure of FE65 WW in the apo form was determined by SAD methods, using SeMet-substituted

protein and data collected on beamline X12C at the National Synchrotron Light Source (NSLS), Brookhaven National Laboratory, Long Island, NY. The structure of the FE65 WW–Mn10 complex was solved by MAD methods, using SeMet-substituted FE65 WW protein and data collected at NSLS, beamline X12B. Initial SAD and MAD phases were calculated using SHELX,⁵⁰ followed by several cycles of model building and refinement using COOT,⁵¹ REFMAC5,⁵² and ARP/wARP.⁵³ The FE65 WW–Mn9 structure was determined by molecular replacement using MOLREP,⁵⁴ and the FE65 WW–Mn10 complex as a search model. PROCHECK⁵⁵ was used for analysis and validation of the refined structures (Table 1).

Isothermal titration calorimetry

Binding constants of FE65 WW for the synthetic peptides Mn9 and Mn10 were measured using a VP-ITC microcalorimeter (MicroCal). Briefly, 0.71 mM Mn9 and 0.74 mM Mn10 were titrated against 63.9 µM FE65 WW protein in 20 mM Tris–HCl (pH 7.6), 150 mM NaCl at 23 °C. Titration curves were analyzed using the program ORIGIN 5.0 (OriginLab). Protein and peptide concentrations were determined by quantitative amino acid analysis on an ABI 420A derivatizer/analyzer and an ABI 103A separation system (Applied Biosystems).

Protein Data Bank accession codes

The atomic coordinates and structure factors of FE65 WW in the apo form and bound to Mena peptides Mn9 and Mn10 have been deposited in the Protein Data Bank with accession codes 2IDH, 2OEI, and 2HO2, respectively.

Acknowledgements

We thank the staff at NSLS for assistance with diffraction data collection. This work was supported by grants AG021964, DK062162, and GM065520 from the National Institutes of Health, DAMD170210300, DAMD170310563, W81XWH0510622, and W81XWH0710178 from the US Department of Defense, and the Temple Discovery Award TLL035927 from the Alzheimer's Association to J.A.A.L.

References

- King, G. D. & Turner, R. S. (2004). Adaptor protein interactions: modulators of amyloid precursor protein metabolism and Alzheimer's disease risk? *Expt. Neurol.* **185**, 208–219.
- Zambrano, N., Bruni, P., Minopoli, G., Mosca, R., Molino, D., Russo, C. *et al.* (2001). The beta-amyloid precursor protein APP is tyrosine-phosphorylated in cells expressing a constitutively active form of the Abl protooncogene. *J. Biol. Chem.* **276**, 19787–19792.
- Ermeikova, K. S., Zambrano, N., Linn, H., Minopoli, G., Gertler, F., Russo, T. & Sudol, M. (1997). The WW domain of neural protein FE65 interacts with proline-rich motifs in Mena, the mammalian homolog of *Drosophila* enabled. *J. Biol. Chem.* **272**, 32869–32877.

4. Masin, M., Kerschensteiner, D., Dumke, K., Rubio, M. E. & Soto, F. (2006). Fe65 interacts with P2X2 subunits at excitatory synapses and modulates receptor function. *J. Biol. Chem.* **281**, 4100–4108.
5. Cao, X. & Sudhof, T. C. (2001). A transcriptionally active complex of APP with Fe65 and histone acetyltransferase Tip60. *Science*, **293**, 115–120.
6. Zambrano, N., Minopoli, G., de Candia, P. & Russo, T. (1998). The Fe65 adaptor protein interacts through its PID1 domain with the transcription factor CP2/LSF/LBP1. *J. Biol. Chem.* **273**, 20128–20133.
7. Trommsdorff, M., Borg, J. P., Margolis, B. & Herz, J. (1998). Interaction of cytosolic adaptor proteins with neuronal apolipoprotein E receptors and the amyloid precursor protein. *J. Biol. Chem.* **273**, 33556–33560.
8. Hoe, H. S., Magill, L. A., Guenette, S., Fu, Z., Vicini, S. & Rebeck, G. W. (2006). Fe65 interaction with the ApoE receptor ApoEr2. *J. Biol. Chem.* **281**, 24521–24530.
9. Borg, J. P., Ooi, J., Levy, E. & Margolis, B. (1996). The phosphotyrosine interaction domains of X11 and FE65 bind to distinct sites on the YENPTY motif of amyloid precursor protein. *Mol. Cell. Biol.* **16**, 6229–6241.
10. Sabo, S. L., Ikin, A. F., Buxbaum, J. D. & Greengard, P. (2001). The Alzheimer amyloid precursor protein (APP) and FE65, an APP-binding protein, regulate cell movement. *J. Cell Biol.* **153**, 1403–1414.
11. Sabo, S. L., Ikin, A. F., Buxbaum, J. D. & Greengard, P. (2003). The amyloid precursor protein and its regulatory protein, FE65, in growth cones and synapses *in vitro* and *in vivo*. *J. Neurosci.* **23**, 5407–5415.
12. Wang, B., Hu, Q., Hearn, M. G., Shimizu, K., Ware, C. B., Liggitt, D. H. *et al.* (2004). Isoform-specific knockout of FE65 leads to impaired learning and memory. *J. Neurosci. Res.* **75**, 12–24.
13. Sabo, S. L., Lanier, L. M., Ikin, A. F., Khorkova, O., Sahasrabudhe, S., Greengard, P. & Buxbaum, J. D. (1999). Regulation of beta-amyloid secretion by FE65, an amyloid protein precursor-binding protein. *J. Biol. Chem.* **274**, 7952–7957.
14. Santiard-Baron, D., Langui, D., Delehedde, M., Delatour, B., Schombert, B., Touchet, N. *et al.* (2005). Expression of human FE65 in amyloid precursor protein transgenic mice is associated with a reduction in beta-amyloid load. *J. Neurochem.* **93**, 330–338.
15. Pietrzik, C. U., Yoon, I. S., Jaeger, S., Busse, T., Weggen, S. & Koo, E. H. (2004). FE65 constitutes the functional link between the low-density lipoprotein receptor-related protein and the amyloid precursor protein. *J. Neurosci.* **24**, 4259–4265.
16. Kimberly, W. T., Zheng, J. B., Guenette, S. Y. & Selkoe, D. J. (2001). The intracellular domain of the beta-amyloid precursor protein is stabilized by Fe65 and translocates to the nucleus in a Notch-like manner. *J. Biol. Chem.* **276**, 40288–40292.
17. Cao, X. & Sudhof, T. C. (2004). Dissection of amyloid-beta precursor protein-dependent transcriptional transactivation. *J. Biol. Chem.* **279**, 24601–24611.
18. Telese, F., Bruni, P., Donizetti, A., Gianni, D., D'Ambrosio, C., Scaloni, A. *et al.* (2005). Transcription regulation by the adaptor protein Fe65 and the nucleosome assembly factor SET. *EMBO Rep.* **6**, 77–82.
19. Hass, M. R. & Yankner, B. A. (2005). A gamma-secretase-independent mechanism of signal transduction by the amyloid precursor protein. *J. Biol. Chem.* **280**, 36895–36904.
20. Sumioka, A., Nagaishi, S., Yoshida, T., Lin, A., Miura, M. & Suzuki, T. (2005). Role of 14-3-3gamma in FE65-dependent gene transactivation mediated by the amyloid beta-protein precursor cytoplasmic fragment. *J. Biol. Chem.* **280**, 42364–42374.
21. Nakaya, T. & Suzuki, T. (2006). Role of APP phosphorylation in FE65-dependent gene transactivation mediated by AICD. *Genes Cells*, **11**, 633–645.
22. Yang, Z., Cool, B. H., Martin, G. M. & Hu, Q. (2006). A dominant role for FE65 (APBB1) in nuclear signaling. *J. Biol. Chem.* **281**, 4207–4214.
23. Guenette, S., Chang, Y., Hiesberger, T., Richardson, J. A., Eckman, C. B., Eckman, E. A. *et al.* (2006). Essential roles for the FE65 amyloid precursor protein-interacting proteins in brain development. *EMBO J.* **25**, 420–431.
24. Gertler, F. B., Niebuhr, K., Reinhard, M., Wehland, J. & Soriano, P. (1996). Mena, a relative of VASP and Drosophila Enabled, is implicated in the control of microfilament dynamics. *Cell*, **87**, 227–239.
25. Lanier, L. M., Gates, M. A., Witke, W., Menzies, A. S., Wehman, A. M., Macklis, J. D. *et al.* (1999). Mena is required for neurulation and commissure formation. *Neuron*, **22**, 313–325.
26. Goh, K. L., Cai, L., Cepko, C. L. & Gertler, F. B. (2002). Ena/VASP proteins regulate cortical neuronal positioning. *Curr. Biol.* **12**, 565–569.
27. Lebrand, C., Dent, E. W., Strasser, G. A., Lanier, L. M., Krause, M., Svitkina, T. M. *et al.* (2004). Critical role of Ena/VASP proteins for filopodia formation in neurons and in function downstream of Netrin-1. *Neuron*, **42**, 37–49.
28. Sudol, M. & Hunter, T. (2000). NeW wrinkles for an old domain. *Cell*, **103**, 1001–1004.
29. Zarrinpar, A. & Lim, W. A. (2000). Converging on proline: the mechanism of WW domain peptide recognition. *Nature Struct. Biol.* **7**, 611–613.
30. Macias, M. J., Wiesner, S. & Sudol, M. (2002). WW and SH3 domains, two different scaffolds to recognize proline-rich ligands. *FEBS Letters*, **513**, 30–37.
31. Ball, L. J., Kuhne, R., Schneider-Mergener, J. & Oschkinat, H. (2005). Recognition of proline-rich motifs by protein-protein-interaction domains. *Angew. Chem. Int. Ed.* **44**, 2852–2869.
32. Ingham, R. J., Colwill, K., Howard, C., Dettwiler, S., Lim, C. S., Yu, J. *et al.* (2005). WW domains provide a platform for the assembly of multiprotein networks. *Mol. Cell. Biol.* **25**, 7092–7106.
33. Pires, J. R., Parthier, C., Aido-Machado, R., Wiedemann, U., Otte, L., Bohm, G. *et al.* (2005). Structural basis for APPTPPPLPP peptide recognition by the FBP11WW1 domain. *J. Mol. Biol.* **348**, 399–408.
34. Kato, Y., Miyakawa, T., Kurita, J. & Tanokura, M. (2006). Structure of FBP11 WW1-PL ligand complex reveals the mechanism of proline-rich ligand recognition by group II/III WW domains. *J. Biol. Chem.* **281**, 40321–40329.
35. Kato, Y., Nagata, K., Takahashi, M., Lian, L., Herrero, J. J., Sudol, M. & Tanokura, M. (2004). Common mechanism of ligand recognition by group II/III WW domains: redefining their functional classification. *J. Biol. Chem.* **279**, 31833–31841.
36. Huang, X., Poy, F., Zhang, R., Joachimiak, A., Sudol, M. & Eck, M. J. (2000). Structure of a WW domain containing fragment of dystrophin in complex with beta-dystroglycan. *Nature Struct. Biol.* **7**, 634–638.
37. Verdecia, M. A., Bowman, M. E., Lu, K. P., Hunter, T. & Noel, J. P. (2000). Structural basis for phosphoserine-proline recognition by group IV WW domains. *Nature Struct. Biol.* **7**, 639–643.
38. Macias, M. J., Hyvonen, M., Baraldi, E., Schultz, J., Sudol, M., Saraste, M. & Oschkinat, H. (1996). Structure of the WW domain of a kinase-associated

- protein complexed with a proline-rich peptide. *Nature*, **382**, 646–649.
39. Kanelis, V., Rotin, D. & Forman-Kay, J. D. (2001). Solution structure of a Nedd4 WW domain-ENaC peptide complex. *Nature Struct. Biol.* **8**, 407–412.
 40. Kanelis, V., Bruce, M. C., Skrynnikov, N. R., Rotin, D. & Forman-Kay, J. D. (2006). Structural determinants for high-affinity binding in a Nedd4 WW3* domain-Comm PY motif complex. *Structure*, **14**, 543–553.
 41. Chong, P. A., Lin, H., Wrana, J. L. & Forman-Kay, J. D. (2006). An expanded WW domain recognition motif revealed by the interaction between Smad7 and the E3 ubiquitin ligase Smurf2. *J. Biol. Chem.* **281**, 17069–17075.
 42. Feng, S., Chen, J. K., Yu, H., Simon, J. A. & Schreiber, S. L. (1994). Two binding orientations for peptides to the Src SH3 domain: development of a general model for SH3-ligand interactions. *Science*, **266**, 1241–1247.
 43. Lim, W. A., Richards, F. M. & Fox, R. O. (1994). Structural determinants of peptide-binding orientation and of sequence specificity in SH3 domains. *Nature*, **372**, 375–379.
 44. Espanel, X. & Sudol, M. (1999). A single point mutation in a group I WW domain shifts its specificity to that of group II WW domains. *J. Biol. Chem.* **274**, 17284–17289.
 45. Nguyen, J. T., Turck, C. W., Cohen, F. E., Zuckermann, R. N. & Lim, W. A. (1998). Exploiting the basis of proline recognition by SH3 and WW domains: design of N-substituted inhibitors. *Science*, **282**, 2088–2092.
 46. Zarrinpar, A., Bhattacharyya, R. P. & Lim, W. A. (2003). The structure and function of proline recognition domains. *Sci. STKE*, **179**, RE8.
 47. Socolich, M., Lockless, S. W., Russ, W. P., Lee, H., Gardner, K. H. & Ranganathan, R. (2005). Evolutionary information for specifying a protein fold. *Nature*, **437**, 512–518.
 48. Russ, W. P., Lowery, D. M., Mishra, P., Yaffe, M. B. & Ranganathan, R. (2005). Natural-like function in artificial WW domains. *Nature*, **437**, 579–583.
 49. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray crystallographic data in oscillation mode. *Methods Enzymol.* **276**, 307–326.
 50. Sheldrick, G. & Schneider, T. (1997). SHELXL: high-resolution refinement. *Methods Enzymol.* **277**, 319–343.
 51. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallog. sect. D*, **60**, 2126–2132.
 52. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallog. sect. D*, **53**, 240–255.
 53. Perrakis, A., Morris, R. & Lamzin, V. S. (1999). Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.* **6**, 458–463.
 54. Vagin, A. A. & Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. *J. Appl. Crystallog.* **30**, 1022–1025.
 55. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283–291.
 56. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res.* **22**, 4673–4680.
 57. Esnouf, R. M. (1997). An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graph. Model.* **15**, 132–134.
 58. Wallace, A. C., Laskowski, R. A. & Thornton, J. M. (1995). Derivation of 3D coordinate templates for searching structural databases: application to Ser-His-Asp catalytic triads in the serine proteinases and lipases. *Protein Eng.* **8**, 127–134.

Edited by I. Wilson

(Received 16 April 2007; received in revised form 19 June 2007; accepted 22 June 2007)
Available online 29 June 2007